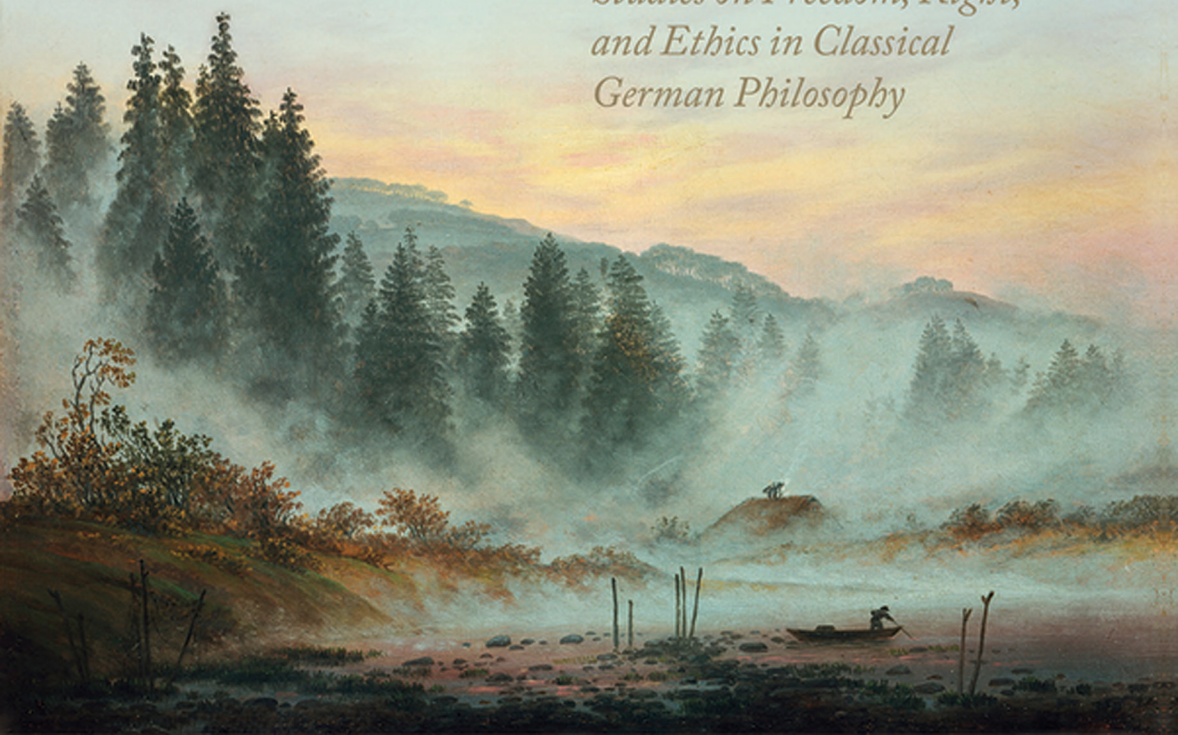


OXFORD

ALLEN W. WOOD

The Free Development of Each

*Studies on Freedom, Right,
and Ethics in Classical
German Philosophy*



The Free Development of Each

The Free Development of Each

*Studies on Freedom, Right, and Ethics
in Classical German Philosophy*

Allen W. Wood

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

© in this volume Allen W. Wood 2014

The moral rights of the author have been asserted

First Edition published in 2014

Impression: 1

All rights reserved. No part of this publication may be reproduced, stored in
a retrieval system, or transmitted, in any form or by any means, without the
prior permission in writing of Oxford University Press, or as expressly permitted
by law, by licence or under terms agreed with the appropriate reprographics
rights organization. Enquiries concerning reproduction outside the scope of the
above should be sent to the Rights Department, Oxford University Press, at the
address above

You must not circulate this work in any other form
and you must impose this same condition on any acquirer

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2013952007

ISBN 978-0-19-968553-0

Printed and bound in Great Britain by
CPI Group (UK) Ltd, Croydon, CRO 4YY

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

To three of my favorite colleagues
David Hills
Debra Satz
and (of course)
Rega Wood

Acknowledgements

Chapter 2 is a revision of “Kant on Practical Reason,” in Sorin Baiasu and Mark Timmons (eds) *Kant on Practical Justification: Interpretive Essays*, Oxford: Oxford University Press, 2012.

Parts of Chapter 3 draw on “Right and Ethics: Arthur Ripstein’s Force and Freedom, in Sari Kisilevsky and Martin Stone (eds), *Freedom and Force: Essays on Kant’s Legal Philosophy*, Oxford: Hart Publishing, forthcoming.

A shorter version of Chapter 4 first appeared as “Kant’s Principles of Publicity” in Paul Formosa, Avery Goldman, and Tatiana Patrone (eds), *Politics and Teleology in Kant*. Cardiff: University of Wales Press, 2014.

Chapter 5 is a revised version of “Herder and Kant on History: Their Enlightenment Faith,” in Lawrence Jorgensen and Samuel Newlands (eds), *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*, Oxford: Oxford University Press, 2009.

Chapter 7 is a slightly revised version of “Fichte’s Absolute Freedom,” which first appeared in *Yearbook for German Idealism*, Fred Rush (ed.), Berlin: Walter de Gruyter, Volume IX, Autumn, 2013.

Chapter 8 is a revised version of the article “Fichte’s Intersubjective I,” of which the final and definitive form, the Version of Record, has been published in *Inquiry* Vol. 49, No. 1, Songsuk Susan Hahn (ed.), February, 2006, copyright Taylor & Francis, pp. 62–79, available online at: <http://www.tandfonline.com/> [Article DOI] (accessed June 20, 2013).

Chapter 9 is a revised version of “Fichtean Themes in Hegel’s Dialectic of Recognition,” in Carlos Rendón (ed.), *Figure and Experience: Lessons about Hegel’s Phenomenology of Mind*, Medellin, forthcoming.

Chapter 10 is a revised version of “Hegel on Responsibility for Actions,” in Arto Laitinen and Konstantin Sandis (eds), *Hegel on Action*, Houndmills: Palgrave Macmillan, 2010.

Chapter 12 is a slightly revised version of “Coercion, manipulation, exploitation,” Michael Weber (ed.) *Manipulation: Moral and Political Issues*. Oxford: Oxford University Press, 2014 (forthcoming).

Contents

List of Abbreviations

xiii

Introduction	1
1 Freedom and Rational Agency	2
2 Freedom and Community	5
3 Rightful Freedom	8
4 Freedom and Coercion	10
1 Moral Worth, Merit, and Acting from Duty	13
1.1 Modern Ethics and Ancient Ethics: Moral Value as a Distinctive <i>Kind</i> of Value	14
1.2 Moral Worth and Other Concepts of Moral Value	15
1.3 Moral Worth is Different from Merit	16
1.4 Moral Praise is Distinct from Esteem	20
1.5 “From Duty”: What Kant Does and Does not Mean by This Phrase	21
1.6 How <i>Kantians</i> Usually Go Wrong Here	23
1.7 The Duty to Act from Duty	24
1.8 The Structure of the Values Morality Cares About	31
1.9 Moral Worth, the Good Will, and the Formula of Universal Law	34
1.10 Innocence, Corruption, and the Place of Morality in Human Life	37
2 Kant on Practical Reason	40
2.1 Reason and Reasons	40
2.2 Reasons and Rational Justification	41
2.3 Practical Reason as Self-constraint by Objective Value	43
2.4 Basic Principles and Intermediate Principles	45
2.5 Instrumental Reason	47
2.6 Prudential Reason	52
2.7 Moral Reason	60
2.8 Practical Good	61
2.9 “Virtue” Theories of Practical Reason	63
2.10 Rational Justification	65
2.11 Mutual Recognition and Community	67
3 The Independence of Right from Ethics	70
3.1 Morals, Right and Ethics	70
3.2 Right is Grounded Solely on External Freedom	72
3.3 Ends and the Rational Structure of Action	74
3.4 Right and Universal Law	78

3.5	Right as Grounded on Humanity	80
3.6	Right as External Constraint, and as Duty	81
3.7	Kantian Right is Sooner Socialist than Libertarian	83
4	The Moral Politician	90
4.1	Politics in Kant's Time—and Ours	90
4.2	Kant and <i>Realpolitik</i>	91
4.3	Right and Ethics	93
4.4	Political Wrongdoing	95
4.5	Can Unjust Rulers be Rightfully Coerced?	97
4.6	The Principles of Publicity	99
4.7	Objections and "Counterexamples"	101
4.8	Examples	105
4.9	Rightful Freedom and Beneficial Consequences	108
4.10	Kant on Lying—and Political Lying	111
4.11	Right and Practicality in Politics	114
4.12	The Moralist Speaks to Practical Politicians	117
5	Herder and Kant on History: Their Enlightenment Faith	119
5.1	Herder and the Enlightenment	119
5.2	Herder's Historical Manifesto	121
5.3	Understanding Each Culture from Within	124
5.4	The Failings of the Present Age	125
5.5	Herder and "Cultural Relativism"	128
5.6	Herder vs Kant	132
5.7	Faith in Historical Progress as a Rational Faith	137
6	Leaving Consequentialism Behind	144
6.1	"Consequentialism" vs "Deontology"	144
6.2	What is Consequentialism?	145
6.3	What is Deontology?	146
6.4	Both Insights Together—in Aristotle	148
6.5	Deontology and the Ends of Action	149
6.6	Consequentialism and Human Flourishing	151
6.7	Ends and Means	152
6.8	The Encompassing Good	154
6.9	Consequentialist Reasoning Always Rests on Deontological Grounds	155
6.10	"The Shadow of Hedonism"	157
6.11	Can Consequentialists Value Actions for their Own Sake?	158
6.12	Must Consequentialists and Deontologists Disagree About What to Do?	162
7	Fichte's Absolute Freedom	164
7.1	Fichte's Philosophical Conversion	164
7.2	What is Absolute Freedom?	165
7.3	The "Essence" or "Original Being" of the I	167
7.4	The I as Will	170
7.5	The Conviction that We are Free	173

7.6	Freedom as a Moral Commitment	177
7.7	Freedom as a Presupposition of Theoretical Reason	181
7.8	The Traditional Problem of Free Will	191
8	Fichte's Intersubjective I	194
8.1	Our Cartesian Habit	194
8.2	"Descartes' Error"	195
8.3	A Transcendental Approach to Intersubjectivity	199
8.4	The Other I as Condition for the I's Individuality	203
8.5	The Summons	205
8.6	"No Thou, No I"	210
9	Fichtean Themes in Hegel's Dialectic of Recognition	214
9.1	Transcendental Philosophy in Fichte and Hegel	214
9.2	Self-consciousness in Hegel's <i>Phenomenology</i>	218
9.3	Desire	219
9.4	Intersubjectivity	220
9.5	Subordination and Co-ordination	222
9.6	Self-certainty and Recognition	223
9.7	Struggle and Domination	224
9.8	Freedom and Mutuality	225
10	Hegel on Responsibility for Actions and Consequences	229
10.1	Hegel on Morality	229
10.2	Imputability in Kant and in Hegel	231
10.3	Responsibility, Purpose and Intention	235
10.4	Intention and Motive	237
10.5	The Right of Objectivity, and Negligence	239
10.6	"Double Effect"	243
10.7	"Moral Luck"	246
10.8	Taking Responsibility	249
11	Marx on Equality	252
11.1	Inequality Today	252
11.2	Marx and Engels on the Meaning of Equality	253
11.3	Equality as a Political Concept	255
11.4	The Defects in Any Equal Standard	257
11.5	Equal Right and the Political State	260
11.6	Class Society	262
11.7	Marx and Stirner	265
11.8	Coming to Terms with Marx	267
12	Coercion, Manipulation, Exploitation	274
12.1	Moralized and Non-moralized Concepts	275
12.2	Coercion	277
12.3	What is an <i>Acceptable</i> Alternative?	279
12.4	Coercion, Wrongfulness, and Responsibility	282

xii CONTENTS

12.5 Manipulation	286
12.6 Advertising—and Manipulation Without a Manipulator	291
12.7 Freedom as Independence of the Will of Others	294
12.8 Exploitation	296
12.9 The Connections	301
Concluding Remarks	303
1 All Should be Left Externally Free to Govern their Own Lives	307
2 The External Freedom of All Must be Restricted Merely so that All May be Free	309
3 Hope When the Future Remains in Doubt	314
<i>Bibliography</i>	317
<i>Index</i>	325

List of Abbreviations

Writings by Kant

- Ak *Immanuel Kants Schriften*. Ausgabe der königlich preussischen Akademie der Wissenschaften (Berlin: W. de Gruyter, 1902). cited by volume:page number in this edition
- Anth *Anthropologie in pragmatischer Hinsicht* (1798), Ak 7
Anthropology from a Pragmatic Point of View, Ca Anthropology History and Education
- CA *Cambridge Edition of the Writings of Immanuel Kant* (New York: Cambridge University Press, 1992) This edition provides marginal Ak volume:page citations
- EF *Zum ewigen Frieden: Ein philosophischer Entwurf* (1795), Ak 8
Toward Perpetual Peace: A Philosophical Project, Ca Practical Philosophy
- G *Grundlegung zur Metaphysik der Sitten* (1785), Ak 4
Groundwork of the Metaphysics of Morals, Ca Practical Philosophy
- I *Idee zu einer allgemeinen Geschichte in weltbürgerlicher Absicht* (1784), Ak 8
Idea toward a Universal History with a Cosmopolitan Aim, Ca Anthropology History and Education
- KpV *Kritik der praktischen Vernunft* (1788), Ak 5
Critique of Practical Reason, Ca Practical Philosophy
- KrV *Kritik der reinen Vernunft* (1781, 1787). Cited by A/B pagination.
Critique of Pure Reason, Ca Critique of Pure Reason
- KU *Kritik der Urteilskraft* (1790), Ak 5
Critique of the Power of Judgment, Ca Critique of the Power of Judgment
- MS *Metaphysik der Sitten* (1797–1798), Ak 6
Metaphysics of Morals, Ca Practical Philosophy
- OD *Was heißt: sich im Denken orientieren?* Ak 8
What does it Mean to Orient Oneself in Thinking? Ca Religion and Rational Theology

- R *Religion innerhalb der Grenzen der bloßen Vernunft* (1794),
Ak 6 *Religion within the Boundaries of Mere Reason*, Ca *Religion and Rational Theology*
- Refl Reflexionen, Ak 15
- RH *Rezensionen von Herders Ideen*, Ak 8
Reviews of Herder's Ideas, Ca *Anthropology, History and Education*
- SF *Der Streit der Fakultäten* (1798), Ak 7
The Conflict of the Faculties, Ca *Writings on Religion and Rational Theology*
- TP *Über den Gemeinspruch: Das mag in der Theorie richtig sein, taugt aber nicht für die Praxis* (1793), Ak 8
On the Common Saying: That may be Correct in Theory but it is of No Use in practice, Ca *Practical Philosophy*
- VE *Vorlesungen über Ethik*, Ak 27
Lectures on Ethics, Ca *Lectures on Ethics*
- VL *Vorlesungen über Logik*, ed. Jäsche, Ak 9
Lectures on Logic, Ca *Lectures on Logic*
- VRL *Über ein vermeintes Recht aus Menschenliebe zu lügen*, Ak 8
On a Supposed Right to Lie from Philanthropy, Ca *Practical Philosophy*

Writings by Herder

- Herder Suphan *Werke*, ed. Bernard Suphan. Berlin: Weidmann, 1877–1913. Cited by volume: page number
- Herder Gaier *Werke*, ed. Ulrich Gaier. Frankfurt am Main: Deutscher Klassiker Verlag, 1985. Cited by volume: page number
- PW Herder: *Philosophical Writings*, ed. Michael Forster. Cambridge: Cambridge University Press, 2002. Cited by page number

Writings by Fichte

- EW Daniel Breazeale (ed.) *Fichte: Early Writings*. Ithaca: Cornell University Press, 1988. Cited by page number
- FGA (1962) *J. G. Fichte-Gesamtausgabe*. Eds Reinhard Lauth and Hans Gliwitzky. Stuttgart-Bad Cannstatt: Friedrich Frommann. Cited

by part/volume:page number or in the case of letters in III, by letter number

- IW Daniel Breazeale (ed. and tr.) *Introductions to the Wissenschaftslehre and Other Writings*. Indianapolis: Hackett, 1994
- SW (1970) *Fichtes Sämmtliche Werke*, ed. Immanuel Hermann Fichte. Berlin: Walter deGruyter. Cited by volume: page number

Individual works will be abbreviated as follows

- ARD *Aphorisms on Religion and Deism*, SW 5
- BHW *On Stimulating and Increasing the Pure Interest in Truth* (1795), SW 8, EW
- BWL *On the Concept of a Wissenschaftslehre* (1793), SW 1, EW
- EE *First Introduction to the Wissenschaftslehre* (1797), SW 1, IW
- GEW *Outline of the Distinctive Character of the Wissenschaftslehre* (1795), SW 1, EW
- GGW *On the Basis of our Belief in a Divine Governance of the World* (1798), SW 8, IW
- GH *Der geschlossene Handelstaat* (1800), SW 3
- GWL *Foundation of the Entire Doctrine of Science* (1794), SW 1
- NR *Foundations of Natural Right* (1796), SW 3, ed. F. Neuhouser, tr. Michael Baur. Cambridge: Cambridge University Press, 2000
- SL *System of Ethics* (1798), SW 4, tr. Daniel Breazeale and Günther Zöller. Cambridge: Cambridge University Press, 2006
- VBG *Some Lectures Concerning the Scholar's Vocation* (1794), SW 6, EW
- ZE *Second Introduction to the Wissenschaftslehre* (1797), SW 1, IW

Writings by Hegel

- Werke G. W. F. Hegel, *Werke*. Theorie Werkausgabe. Frankfurt: Suhrkamp, 1970. Cited by volume
- EG G. W. F. Hegel, *Philosophy of Spirit*, tr. William Wallace and A. V. Miller. Oxford: Oxford University Press, 1977. Cited by paragraph (§) number. Werke 12
- EL G. W. F. Hegel, *The Encyclopedia Logic*, tr. Theodore F. Geraets, Wahl Suchting, and H. S. Harris. Indianapolis: Hackett, 1991. Cited by paragraph (§) number. Werke 10

- NP *Nürnberger Propädeutik*, Werke 4
- PhG G. W. F. Hegel, *Phenomenology of Spirit*, tr. A. V. Miller. Oxford: Oxford University Press, 1975. Cited by paragraph (§) number in this translation. Werke 3
- PR G. W. F. Hegel, *Elements of the Philosophy of Right*, ed. Allen W. Wood, tr. H. B. Nisbet. Cambridge: Cambridge University Press, 1991. Cited by paragraph (§) number; the Preface, by page number in this translation. Werke 7
- VG *Die Vernunft in der Geschichte*, ed. J. Hoffmeister. Hamburg: Meiner Verlag, 1955. *Lectures on the Philosophy of World History: Introduction*, tr. H. B. Nisbet. Cambridge: Cambridge University Press, 1975
- VPR *Vorlesungen über Rechtsphilosophie*, ed. Karl-Heinz Ilting. Stuttgart: Frommann Verlag, 1974. Cited by volume:page
- VPR17 *Die Philosophie des Rechts: Die Mitschriften Wannemann* (Heidelberg 1817–1818) *und Homeyer* (Berlin 1818–1819), ed. Karl-Heinz Ilting. Stuttgart: Klett-Kotta, 1983
- VPR19 *Philosophie des Rechts: Die Vorlesung von 1819/1820*, ed. Dieter Henrich. Frankfurt: Suhrkamp, 1983

Writings by Marx

- Capital Marx, *Capital*. tr. Ben Fowkes and David Fernbach. New York: Vintage, 1977–1981. Cited by volume: page
- CW *Marx Engels Collected Works*. New York: International Publishers, 1975. Cited by volume:page

Introduction

The essays in this collection were all written, originally for various and sundry occasions, between 2005 and 2012. I publish them together now basically for two reasons. First, they span the period in the history of philosophy—German philosophy in the late eighteenth and nineteenth centuries, and the figures in that period, that most interest me: Kant, Herder, Fichte, Hegel, Marx. Second, they have a certain thematic unity, dealing with *freedom*—a theme which, more than any other, is the chief theme of German philosophy during that period. Although the chapters are presented more or less in chronological order, in them I do not attempt to trace any systematic development over time of the theme of freedom, or of anything else. But there is a philosophical viewpoint, focusing on the theme of freedom, which emerges from the historical discussion. This, as much as anything, unifies the collection. So the best way of introducing these chapters is to expound it. I do so offering little by way of argument. I hope that it will receive greater clarification and defense in the course of the book. In my concluding remarks I offer some applications, in the spirit of this introduction.

In seeing the thinkers in this tradition as embodying a certain fundamental viewpoint, I am far from denying that there is disagreement between them. Indeed, I try to show how Fichte's concept of absolute freedom differs subtly from Kant's, as well as how Fichte develops the theme of intersubjectivity in a new and creative direction. I also try to show how Kant and Herder differ concerning the correct way of understanding human nature, cultural difference and history; and also how Hegel's theory of imputation differs from Kant's, and deepens it. And, of course, I emphasize the radical difference between Marx's treatment of themes having to do with right and the way these themes were treated by the thinkers earlier in the tradition.

At the same time, in treating the German philosophers of this period whose thought interests me, I always tend to emphasize continuities and agreements

rather than squabbles and the differences. I think it is both shortsighted and wrongheaded to treat these thinkers as though the fundamental issue is whether we should choose Hegel over Kant, or defend Kant against Hegel, or even champion Marx over against the entire later German idealist tradition, trying to show that he has rendered the entire classical German philosophical tradition obsolete (a dogmatic sectarian attitude that is not as fashionable now as it once was). Instead, I think that despite the controversies within this tradition, there is something unified and important in it, when it comes to themes of freedom, right, ethics, humanity, community, and history, which sets the classical German tradition apart from other strands in modern philosophy.

I would also never claim that the viewpoint I draw from these thinkers—especially the conclusions about present-day moral and political issues—is directly present in all of them. I nonetheless do think a careful study of the tradition, and the application of its insights to present-day moral and political reality, justifies the viewpoint. Even the greatest philosophers often hesitate to draw consequent conclusions from their insights, especially when these conclusions would seem to them dangerously radical or too much opposed to the customs and ways of thinking they see around them. It often takes time for the true implications of an insight, or its correct and consistent application to social reality, to become clear. Consequent thinking is often resisted, and rationalizations of even the most abominable social attitudes and practices persist for generations even after clear thinking has exposed them as indefensible. That is my explanation for the various ways in which these historical philosophers might disagree with the moral and political conclusions I want to draw from them.

In these chapters my first obligation is to interpret these philosophers accurately in their own terms, and mostly in relation to the issues of their own time. That is why I formulate the viewpoint central to the following essays—and its application to present-day morality and politics—mainly in this introduction, and also (even more pointedly) in my concluding remarks. I now present this viewpoint by formulating four main theses about freedom:

1 Freedom and Rational Agency: *Freedom, as the Capacity to Govern Oneself Rationally, is Also the Capacity to be the Ultimate Ground of One's Own Identity as a Rational Agent*

Fichte on absolute freedom. Fichte regarded freedom as the foundation of his entire philosophical system. “Absolute freedom,” which is the focus of Chapter 7, involves being situated in a world, both material and social, as an embodied agent

standing essentially in relation with other agents. Absolute freedom is also essentially situated in time—freedom means having an open future, consisting of real options between which one can choose for reasons. It is the human vocation, Fichte thought, to address the situation offered by one's absolute freedom, to make oneself into whatever one can ever become. It is equally the task of each and every human being to fulfill this vocation, and the recognition that this is our vocation is, for Fichte, the achievement of the modern age. In fact, Fichte thought it was the foundation of nothing less than a whole new way for human beings to be in the world. As he put it in one of his lectures:

With the discovery of this philosophy an entirely new epoch in the history of the human species has begun—or, if one prefers, an entirely new and different human species has arisen, one for which all previous forms of human nature and activity on earth are no more than preparatory, if they retain any value at all. This is the philosophy to which our age summons us all and which we can all take a hand in developing just as soon as we have a desire to do so (GA II 3:335).

Kant on acting from duty. The notion of a self-governing rational agent, however, is something that Fichte thought he had gotten from Kant. This Kantian conception is explored in Chapters 1 and 2. The conception of practical reason is basic to all practical philosophy in the Kantian tradition. In Chapter 2, I examine Kant's chief thematic presentation of this conception, in the Second Section of the *Groundwork for the Metaphysics of Morals*, and attempt to make of it a single, coherent, and self-consistent set of doctrines about the norms of reason and their applicability to actions. The conception of freedom as rational self-government is one that takes freedom to consist most fundamentally in the capacity of reason to constrain ourselves, to *act from duty*, as Kant famously presents it in the opening pages of the First Section of *Groundwork*. To Kant, the human moral situation is deeply paradoxical and troubling. I attempt to present his views on this point sympathetically in Chapter 1.

I have long believed that this famous discussion has been badly misunderstood, not only by Kant's critics, who typically turn it into a caricature that is all too easily dismissed, but even by most of Kant's defenders and sympathizers, who (in my opinion) miss the main point of the discussion. This point has to do with the centrality to morality—and to the function of morality in human life—of the importance, and also the difficulty for us human beings, of making ourselves do things because moral reason requires them. Morality is therefore most itself not when it is easy or natural for us, but when it is difficult; and it is our human condition to act with greatest freedom when what we do does not come naturally or easily but only when it requires self-overcoming to do it.

This was to be a point of controversy in the early reception of Kantian ethics, especially in the form of Schiller's complaint (echoed in the early theological writings of Hegel, and often repeated since) that true freedom, and true moral virtue, would involve a harmony of inclination with reason, rather than a self-constraint by reason against the resistance of natural desire. (A recent discussion of this theme in Kant and several of his followers and critics is to be found in Stern, 2012.) Kant's thesis that morality, and the freedom it represents, is centrally about duty (about rational self-coercion and self-constraint) is not a thesis about the normative force of practical principles. He acknowledged that "holy" beings could follow rational principles with no need for constraint. Kant's is rather a darker thesis about our human nature—perhaps even our nature as "civilized" human beings (in a post-agricultural or even a modern capitalist society)—and the role that morality must play in human life as a consequence of our flawed or corrupted *social* nature. It seems clear to me, at least, that regarding our present society, Kant is right. The brighter, antinomian vision championed by Schiller may someday be suited to life in a better world than the one we live in, but Kant is right about *our* world.

Free action for its own sake. *Action* is, by its concept, something done for *reasons*. It is also always something done for an *end*—an action is, by its concept, something chosen in order to produce some result or consequence in the world. But the evaluation of states of affairs as the possible outcomes of actions ought to depend on the evaluation of the actions that set ends, not the other way around. And the reasons that bear on the evaluation of actions are many and heterogeneous. Some have to do with the desirability of consequences, but others have to do with principles to be followed, or with the enactment of the agent's self-conception. Using Fichte as well as Aristotle to make the point, Chapter 6 argues that the deontological conception of practical value, which gives primacy to choosing actions for their own sake, should have priority over the choice of actions as means to ends external to them. Consequentialist reasoning has its proper place in deciding what to do, but that place is secondary to, and consequent upon, the value or disvalue of actions for their own sake.

Action and imputability. As free agents, we care about the consequences of actions because they belong to us, or in traditional terms, because they may be *imputed* to us. Kant has one theory of imputation, but a deeper theory, I believe, was developed by Hegel. It is customary to see Hegel as using "morality" (*Moralität*) as only a nickname for Kantian (or Kantian–Fichtean) moral philosophy, contrasting it with Hegel's own viewpoint, under the title "ethical life" (*Sittlichkeit*). But I think this way of viewing the matter is fundamentally wrong. In Chapter 10, I argue that the Kantian conception of a rationally self-governing free agent was not the only version to be found in the classical German tradition: it was already significantly

modified by Fichte, and emerged in a more profound form in Hegel's conception of moral subjectivity. I believe Hegel has a distinctive and positive concept of *moral subjectivity* (his term for it), which is both different from Kant's and capable of being viewed as a valuable corrective to it, or even as a "friendly amendment."

2 Freedom and Community: *The Freedom of Individual Subjects is Grounded Transcendentally in a Relation to Other Free Subjects, and to Distinctive Forms of Community Between Free Subjects*

Morality should *not* be seen as a coercive or manipulative device by which society controls its members, forcing them to comply with the requirements it imposes on them. In saying this, however, I am not disagreeing with the psychology and sociology of Nietzsche, Freud, or Foucault, who convincingly portray the historical reality of morality in these nasty terms. They may be right that this is what "morality" in fact has been, and still is, in many places, or even in all places. The point is rather that this is not what moral principles, duties or reasons, as we live them, when we do so as rational beings—as we should live them—are about, or *should be* about. Morality should have to do solely with the way we rationally govern ourselves. Morality, though it involves constraint (self-constraint based on reasons), and in fact just because it involves these, is really about *freedom*.

It might seem as though the concept of freedom as rational individual self-government, and a view that makes this freedom central to what is valuable about human life, would result in a form of "individualism." And it does, in certain ways: for instance, by insisting on individual responsibility in treating issues of imputability, or by defending the claims of individual freedom in the context of right (which will be discussed presently under the heading of my third main thesis). But freedom as rational self-government is not individualistic if "individualism" is understood as a polar opposite to the thesis that human community is central to what is important and valuable in human life. It is central to the insights of the German classical tradition, as I understand it, to reject the view that the value of individual freedom must exist in some kind of tension with values associated with community. On the contrary: *those who see the value of freedom and the value of community as in tension cannot properly understand either value; a society that treats these values as rivals, or thinks it must trade one value away in order to make any place at all for the other, will never fully achieve either.* This is a theme that runs through classical German philosophy. It reaches fruition in different ways in Fichte's conception of the self-sufficiency of reason, Hegel's conception of ethical life, and Marx's conception of social individuality.

The intersubjectivity of the I. Underlying the unity of the values of individual freedom and human community in the classical German tradition is an ontological or transcendental thesis about the nature of mind, reason, and selfhood. These are, namely, constituted by relationships between individual minds and agents. In Kant, this aspect of my second thesis was anticipated in the importance he gave to free communication in making possible the very existence of reason, and in the way Kant regarded thinking from the standpoint of others as indispensable to thinking for oneself. The way that Kant's theory of reason anticipated these ideas, without quite reaching them, is discussed briefly at the end of Chapter 2.

It was Fichte, however, who took the decisive step of arguing that intersubjectivity is a transcendental condition for the possibility of being a free self at all. This argument in Fichte, and the way that it overcomes the Cartesian tradition in the philosophy of mind, is the focus of Chapter 8. Human thoughts are never merely the product of one individual, but occur essentially in the social process of communication, or at least in the preparation for or aftermath of this process. Ignoring this point is responsible for some of the shortcomings of self-styled "materialism" or "naturalism" in the philosophy of mind. These views think of themselves as anti-Cartesian, but in fact succumb to a Cartesian error even more basic than that of the substance dualism on which they like to dwell. For rejecting substance dualism still leaves most philosophers and neuroscientists with Descartes' notion that if thoughts cannot be metaphorically "inside us"—in a nonspatial, immaterial, inner place—then they may still be located literally *inside us*, inside our heads, as states or processes going on in our brains. This merely repeats the deeper Cartesian error.

No doubt without human brains we could not think, just as without human neural systems we could not vote, or perform Shakespeare's plays, or play baseball. But it would be a fundamental error to infer that any of these activities consists only in what happens within our skulls, as if we did not need voting machines, theaters, baseball diamonds, and—most importantly—*other people* in order to engage in these activities. Fichte's thesis is that this is just as true of self-conscious thinking as it is of these other social activities. Fichte argues that the same is equally true of self-conscious thought and action. These too require mutual relations between free beings if they are to occur—and they also presuppose absolute freedom if they are to be properly understood. The transcendental possibility of the free, self-governing individuality discussed in my first thesis, therefore, is grounded in human communication, the basic form of human community. Human individuals, as rational agents, can come to be only in a community with others. We must be educated or brought up (*erzogen*) to be rational individuals, through a process of communication. Basic to this process is what Fichte calls the "summons"

(*Aufforderung*)—the representation, communicated to one rational being by another—of a possible way of acting freely, which offers us for the first time something like a ground or reason for action. Consciousness, thinking, acting for reasons, all these are possible only for free human selves, embodied agents in relations of rational communication with other free, embodied human thinkers and agents.

Quite an accessible statement of the gist of Fichte's position on the intersubjectivity of mind is given by the sociologist of philosophy, Randall Collins:

Naturally one will not find ideas amid the machinery of the brain, or the computer . . . Ideas are not thing-like at all, except insofar as we represent them in symbols written on materials such as paper, but are first of all communication, which is to say interaction among bodily humans. To enter into the physical brain (or inside the computer) is precisely the wrong way to perceive ideas; for ideas are in the process of communication between one thinker and another, and we perceive the ideas of another brain only by having them communicated to us. It is the same with oneself: one perceives one's own ideas only insofar as one is in a communicative mode. There is no thinking except as the aftermath or preparation of communication (Collins, 1998, p. 2).

Individualism and collectivism. The most radical form in which free individuality is united with community is, as I have already mentioned, to be found in the philosophy of Marx. The unity of individualism with collectivism in Marx's thought emerges in Chapter 11 as basic to his understanding of the possibility of human emancipation in a future classless society. Marx had no determinate conception of such a society, however. Probably the single most common error about Marx is that he had some conception of a "communist egalitarian utopia"—a conception that was then attempted in practice in eastern Europe and elsewhere, with deplorable results, for which Marx now may be blamed. On the contrary: Marx was *not* an egalitarian, and he had *no* definite conception of what future society ought to be like, except by contrast with what he saw in the present capitalist society. He left these matters to what he saw as a working class movement still very much in the process of formation, about which he had very high hopes. Marx's life work, however, was to study really existing capitalism, whose foundations, even despite the passage of time since he wrote, Marx probably understood better than anyone who has ever lived.

The same fundamental values that moved Marx—freedom and community, regarded as mutually requiring one another—were earlier present in the tradition, in the form of Herder's conception of *Humanität*, which was the focus of his rational hope for a more human historical future—though he too, like Kant and Fichte, and above all Marx, had no determinate conception of what such a society would be like. All had only a principled hope, grounding action in a progressive direction. These hopes—and the rational faith that goes with them—are explored in Chapter 5.

3 Rightful Freedom: *The Basic Social Condition for the Existence of Free, Self-governing Human Agents is Right (Recht): The Freedom From Having One's Choices Constrained by the Choices of Others, Insofar as all Enjoy External Freedom Equally According to Universal Laws*

The foundation of all genuine human community is the mutual recognition of each person by every other as a free and equal person. The Kantian formulation of this is found in the concept of *right*. The focus of Chapter 3 is the articulation of the reasons why we must unconditionally value what Kant calls a *Rechtzustand*—a “state” (or condition) of right. These reasons, I argue, are wholly independent of *ethics*—of the Kantian categorical imperative, whose function is to be the standard according to which rational agents govern their own lives. It is for Kant the essential function of a law-governed civil society or political state to achieve, preserve, and perfect a condition of right, in which each person is externally free to the extent that the freedom of each is compatible with a like freedom for all others according to universal laws.

External freedom and recognition. Fichte was to derive this idea of external freedom from the basic conditions for being a rational, self-governing human individual. Individuality arises only through a process of mutual recognition (*Anerkennung*), constituting a “relation of right” (NR 3:41–56). The same basic idea was defended and creatively developed by Hegel, in the form of his famous “master–servant” dialectic. Fichte and Hegel develop different, and complementary, arguments for essentially the same (fundamentally Kantian) thesis: namely, that recognition, personhood, external freedom—these can exist only to the extent that recognition is reciprocal and equal. In short, no one can be free unless all are free. As Fichte once put it: “The only person who is himself free is that person who wishes to liberate everyone around him and who . . . really does so” (VBG 6:309). The Fichtean background of Hegel’s theory of recognition, and the basic continuity between their thought on this topic, are explored in Chapter 9.

External freedom, the kind required by Kantian or Fichtean right (and by Hegel’s abstract right) is the freedom to make choices for oneself, unconstrained by the will of anyone else. It is in that way essentially a social relation—it consists not in the absence of obstacles to doing what one wants, but only in the absence of a certain kind of external determinant to this—namely, the choice of another, constraining, overriding, or subverting our own choice. And it is a freedom that is not only consistent with our being coerced in certain ways, but even requiring this as a condition that others may have the same freedom.

Rightful power over others. For Kant and Fichte, it was the function of the political state to perform the acts of coercion needed to protect rightful freedom. For Kant, the coercion involved in political life requires *rightful authority*—the right to command others to do or refrain from certain actions, and to compel obedience to these commands. This authority, in his view, can itself be exercised rightfully, or it can be abused, and the conditions for its rightful exercise are in that case not fulfilled. Because Kant thought that the hierarchy of these authority relations had to come to an end in the will of a supreme commander—and also because he himself always lived in a despotic monarchy, in which there could be no thought of limiting the power of the supreme commander—it was important to Kant that those who exercise political authority should be *morally* bound (self-constrained) by principles of right, even when no one could be in a position to enforce these through external coercion. These are the constraints that “moral politicians”—politicians who comply with the indispensable conditions of the rightful exercise of political authority—must impose on their own actions. If they do not, then the rightful condition is reduced to one of mere despotism.

Kant’s fullest account of these constraints is found in the appendix to *Toward Perpetual Peace*, in the form of two “principles of publicity,” by which he thinks the holders of political power may know whether their actions are consistent with right. In Chapter 4, I attempt to explicate these principles and Kant’s arguments for them, and to offer a limited defense of Kant’s theory of the rightful boundaries of political power. Although political life in Kant’s time and place was very different in some ways from the form it takes for us, we still may have something to learn from Kant’s account of the rightful limits of political authority.

The Kantian concept of right is the concept of a kind of coercion that may be legitimately exercised on free human beings, as a condition of preserving their freedom according to universal law. As we see towards the end of Chapter 3, these constraints must include constraining the actions of the wealthy and powerful in order to preserve the freedom of those with less wealth and power. Kant’s own thoughts on this topic, as we see towards the end of that chapter, were in one way robust, and pointed towards economic redistribution as a necessary means of preserving rightful freedom. But in other ways, they were limited by the political conditions of his time. Fichte was more farsighted in this respect, and it was clear to him that radical changes not only in the political structure of society but also in its economic relations would be required for the protection of the rightful freedom of all.

Still more radical was Marx, who held that true freedom was incompatible with any social system conceived in terms of rightful coercion—or indeed, in any terms of justice, equality, or other notions grounded on right—which he took to be an

essentially bourgeois notion, one that in practice serves bourgeois class interests. Chapter 11 explores Marx's views on this matter, contrasting them with the views of earlier thinkers in the classical German tradition, but not attempting to resolve the problems posed by the tension. To do that would in effect be to decide what would be the way forward for human society in the modern era, and whether the concept of right has a place in that higher future. I hope my readers may forgive my modesty in not pretending to be able to decide these questions. It is immodest enough if I claim, as I do, that the thoughts of the philosophers treated in this volume are vital to our understanding of them, and to any way forward in answering them, either in theory or in practice.

4 Freedom and Coercion: *Freedom and Coercion Are Not Always Opposites*

"Freedom" is a word with many uses, but often it is meant to refer to something worth valuing. Two such notions of freedom are freedom as successful rational self-government on the part of an individual, and freedom as the condition of a society in which everyone enjoys the rightful freedom to lead their lives independently of the arbitrary power of others. If freedom means either of these things, then I think we can see that there are some good reasons for doubting that coercion is always the opposite of freedom. Sometimes it is even part of what freedom consists in.

If Kant (as I present him in Chapters 1 and 2) is right, then the self-governing freedom of the individual rational agent actually consists, in part, in a certain form of coercion—the self-coercion of acting from duty. Moreover, even that species of freedom to which coercion might seem to be most directly opposed—that external freedom from constraint by another's will, whose protection allegedly consists in a system of right—itself requires coercion—coercion by a just political state—if it is to exist at all. The importance of the concept of coercion to the themes in this book has therefore moved me to end with Chapter 12, in which my aim is not historical but philosophical clarification. In it, I formulate a concept of coercion, and relate it to two other concepts also closely allied to our understanding of freedom—manipulation and exploitation.

When it comes to self-coercion and also to rightful state coercion, the enemy of freedom is never coercion as such, but only *wrongful* coercion or coercion *contrary to reason*. Issues about freedom should never be merely about whether coercion is present, or whether freedom is restricted. They are about *who* is being coerced, and *by whom*, whose *freedom* is being expanded or restricted, and whether the freedom that is gained or restricted is *rightful* freedom.

Lovers of freedom are, of course, also by and large lovers of tolerance and permissiveness. Being tolerant and permissive, especially toward others, but even at times toward oneself, are virtues. But as virtues, tolerance and permissiveness always have limits—there is always the intolerable and the impermissible. *Unlimited* tolerance and permissiveness are not virtues; in the end, they are even enemies of freedom. Freedom, therefore, is not wholly opposed to coercion. Freedom even *consists*, in part, in certain kinds of coercion. Moral autonomy consists in *rational* self-coercion, civil freedom in *rightful* external social (state) coercion.

There are many errors, rationalizations, and wrong notions that follow from the mistaken assumption that freedom and coercion are simply opposites. One is the resistance to, or principled rejection of, the Rousseauian–Kantian insight that for beings like ourselves, true inner freedom does not consist in the absence of all coercion (free rein to our impulses, inclinations, habits, opinions, and whims); it consists instead in self-coercion, rational self-government. For rational and social beings like ourselves, it is always a self-deceived wish that we might reach a condition (perhaps thought of as true moral health or virtue) in which self-constraint would be superfluous. A second deplorable tendency is that of taking rational argument itself, when it is directed at you, as a form of wrongful external coercion, an infringement of your freedom, since it might occasion your being forced (through self-coercion) to change your opinions (so as to accord with the evidence, rather than with your prejudices) or even to change your conduct (perhaps at the cost of no longer indulging some of your treasured impulses, habits, and whims).

Yet another error is the representation of rightful state coercion as always an enemy of freedom, and the state itself, therefore, as something necessarily wrongful and evil. Of course, those who do this are not thinking of the state coercion that protects them, their privileges, and their property. But that is because they see this coercion as no limit on freedom because it is directed only at “others” (“the criminal element,” “the underclass,” or people with an alien ethnicity, with skin darker than their own, or the latest wave of immigrants). State coercion is felt by some to be evil and unjust only when it threatens the freedom to use power, wealth, and privilege to oppress these others. Then we must oppose this coercion in order to protect “freedom”—that is, *our* freedom, or the freedom of the wealthy and privileged whose rule we take for granted—often even at the expense of our own interests, or our own freedom.

Western capitalist society, and especially my own American society, is one characterized by great inequalities. In any such society, by the nature of the case, the greatest threat to rightful freedom is always the wealth and power of the privileged. The chief task of the state in protecting human freedom should always be to use rightful state coercion to limit the freedom of the powerful and privileged

to infringe the rightful freedom of the less privileged and the vulnerable. Political struggles in the modern world are usually fundamentally struggles about whether state power will be used to protect the rightful freedom of all, or instead used to protect the wrongful freedom of the wealthy, powerful, and privileged. Wide social inequality necessarily indicates that these struggles have come out the wrong way, on behalf of the unjust and oppressive freedom of the privileged against the rightful freedom of the majority. What Marx saw clearly is that in societies characterized by class division and class oppression, there is a systematic tendency for state coercion to be used mainly in that way only, and seldom or never on behalf of those who most need it—to such a degree, in fact, that Marx tended to think that the very notions of “right” and “justice” apply in practice only to the use of state power that harmonizes with class oppression. Human freedom would then lie on the side of the revolutionary overthrow of the existing use of state power in the interest of ruling-class wealth and privilege, rather than in using state power to protect “right” and “justice.” The problem with this is that it might once again encourage the notion that freedom and coercion are simply opposites, and that the free society of the future would be one in which there is (or need be) no coercion, whether inner or outer. But Marx himself recognized (contrary to Bakunin and the anarchists) that the proletarian revolution could not involve the immediate abolition of the state, and that in the early stages of communism the “narrow horizon of bourgeois right” could not be crossed immediately, but rather there would have to be some system (though Marx thought it would always be beset by defects and limitations needing to be transcended) of *rightful* distribution of the means of consumption (CW 24:86–7). Marx appears to have followed Fichte in thinking that at some point in the distant historical future “the state falls away as legislative and coercive power” (SL 4:253).

The question posed here is whether human social life can ever do without rightful coercion, or even whether it should try. If freedom is not ultimately opposed to coercion, but rather partly *consists* in coercion of the right kinds, then it would be misguided even to *aspire* to freedom entirely without coercion. That, however, remains a *philosophical* question. No rational person could think that freedom entirely without coercion is possible *now*, either in the case of individual self-government or for a society that respects the rightful freedom of its members.

The approach taken in this book, especially in its early chapters (which are about Kant), assumes that we are asking questions about freedom and coercion (both external state coercion and moral self-constraint) from the standpoint of existing cultural reality. We thus begin in Chapter 1 by considering Kant’s crucial idea (too often misunderstood) that begins the *Groundwork for the Metaphysics of Morals*: the idea that morality, in order to be about freedom, must also be fundamentally about *rational self-constraint*.

Moral Worth, Merit, and Acting from Duty

Many who are acquainted with Kant's moral philosophy have a basically negative attitude towards it. Many others respect it only with serious reservation and qualification. Most of these attitudes can be traced in large part to people's reactions to the familiar opening pages of the *Groundwork*, in which Kant claims that only actions done from duty have moral worth. I have long believed that the meaning of this claim is widely misunderstood, not only by Kant's critics, but (at least in part) by most of his sympathizers and defenders. Kant himself may be in part responsible for the misunderstandings by not anticipating and forestalling them as he should have. Most misunderstandings, however, result from what readers bring to Kant's discussion, leading them to attribute to him a set of attitudes he did not have and claims he never intended to make. The errors result, in my view, from importing too much moral psychology into the examples he discusses and into what he says about them.

This is disastrous if the moral psychology is anti-Kantian, but it is harmful even if too much *Kantian* moral psychology is brought in, because this tends to distract us from the point Kant is making in the opening pages of the *Groundwork*. Even Kantians tend, I believe, to read too much moral psychology into the examples, especially nasty psychology into the examples of dutiful actions not done from duty. They do this, first, because they tend to think (mistakenly, in my view) that when Kant says that an action not done from duty lacks "moral worth" (or its maxim "moral content"), he means to be criticizing that action in comparison to an action done from duty (which has "moral worth" and its maxim moral content). And second, because they want to see these examples as illustrating various Kantian points about moral disposition, virtue, or the duty to act from the motive of duty—none of which, I believe, is Kant's main concern in discussing these examples early in the First Section of the *Groundwork*.

1.1 Modern Ethics and Ancient Ethics: Moral Value as a Distinctive *Kind* of Value

Let's begin with the idea of "moral worth." This idea depends on the thought that although there are many things we value, and even care about from the moral point of view, a special place among them is occupied by morality itself. This is the thought that morality itself has a special value different in kind from all others. Kant not only accepted this thought, but also saw it as crucial to the historical development of morality and moral philosophy. According to Kant, this idea was foreign to the ancients, for whom ethics was the study of the ideal—the best kind of life, in which the ideas of moral good and physical good, morality and happiness, were fused (or confused)—as they also are (in Kant's view, less innocently) in the corrupt attitude of those who take reproaches of prudence for those of conscience (VE 27:248–9, 251). Kant thinks it took the supernatural purity of the Christian ideal to teach people that no human being can achieve the ideal—hence that all ideal systems are in practice impure, accommodating their conceptions of virtue to human weakness and corruption. Moral perfection, he holds, can be adequately conceived of not as an ideal to be achieved, but only as a principle we address to ourselves as an imperative to which we ought to conform (VE 27:251–2; 29:260–1).

Such a modern conception of a morality of principles rather than of ideals led, in turn, Kant holds, to the separation of moral value proper from all others. This took place at first in the form of the notion that there is a human *sense* or *feeling* for moral value, distinct from the sense for pleasure or well-being, that involves "an inner reverence for the law." Kant finds this in the philosophies of Shaftesbury and Hutcheson (VE 27:253; 29:625–6). He thinks the moral sense theorists were mistaken in holding that moral value is known by sense rather than reason, but in his moral psychology he always maintains an important place for moral feeling as an effect of moral reason on our senses (G 4:401n, 442, KpV 5:72–8, MS 6:399–402). This is too seldom appreciated by readers, who react negatively to the opening pages of the *Groundwork*. They think that when he denies moral worth to beneficent actions done from sympathy, he is denying that feeling has any role to play in moral action or motivation. But this is not true.

Kant draws such a distinction between moral and non-moral good when he distinguishes the "good" from "well-being," or the "moral" good from the "physical" or "natural" good (KpV 5:59–62, 112, 122; TP 8:282; Anth 7:276–80). His most characteristic doctrine here is that the genuine goodness of the natural good is conditional on its being combined with the moral good, or, to put it in a terminology used by John Rawls, that the moral and natural goods are "lexically ordered" (G 4:393–4; KpV 5:115–19). This doctrine makes it clear that both are considered

good or desirable from the standpoint of morality—the moral good unconditionally, the non-moral good conditionally on its suitable combination with the moral good. That which has the special value belonging to morality itself, therefore, is by no means the only thing that is to be valued from the moral standpoint.

1.2 Moral Worth and Other Concepts of Moral Value

What is morally good, however, also shows itself in a variety of forms. Two of them discussed early in the *Groundwork* are the unlimited goodness of the good will (G 4: 393), and “the worth of character, which is moral and the highest without any comparison” (G 4:398–9). Neither of these should be confused with yet a third manifestation of the morally valuable that soon becomes the focus of Kant’s discussion: namely, the “moral worth” (*moralische Wert*) attaching to actions that are done from duty (*aus Pflicht*). Kant introduces this concept of moral worth, by means of the concept of duty with which it is associated, which he describes as the way the good will shows itself “under certain subjective limitations and hindrances, which, however, far from concealing [the special value of the good will] and making it unrecognizable, rather elevate it by contrast and let it shine forth all the more brightly” (G 4:397).

It is a crucial feature of Kant’s concept of morality that he thinks the value most central to it shines forth most brightly only under conditions of adversity, in which subjective factors tend to present us with obstacles to our doing the right thing. In that sense, Kant’s concept of morality could be called “anti-naturalistic.” I mean this not in the fashionable sense in which the *natural* is opposed to the supernatural or identified with the objects of natural science, but rather in the sense in which the natural is the same as the easy, the instinctive, the healthy, and the innocent, and the opposite of the *natural* is the uncomfortable, the forced, or that which has been meddled with by human hands. Kant’s central claim, as we will see presently, is that morality is most truly itself precisely when it is not “natural” in this sense—that is, when it is difficult, strained, even artificial, because it is a function of our rational nature, which, according to the view he adopted from Rousseau, is a product of our own activity as the sullied, wretched, socially self-made beings that we are.

Kant uses the term “moral worth” to refer to the kind of moral value that lies right at the heart of morality. This is indicated by the modifiers he often attaches to the term “moral worth”: he speaks of the moral worth of actions as their “inner,” “true” or “authentic” moral worth (G 4:398). Kant’s strategy in the First Section

of the *Groundwork* is to justify his first formula of the moral law, the Formula of Universal Law (FUL), by deriving this formula from the value that is most central to morality. He tries to do this by enlisting the assent of “common rational moral cognition” to his proposition that this central moral value is the true, authentic, or inner worth of actions, the worth they have when they are done from duty.

Before considering the claim that true moral worth belongs only to these actions, it is worthwhile to pause over the concept of this kind of worth, because I think Kant’s readers have shown themselves susceptible to a number of errors about this concept. Some of these errors are hard to avoid, and also contribute mightily to other common misunderstandings of the early pages of the *Groundwork*. One notable feature of the concept of moral worth is that it is the concept of a certain *kind* of value, and *not* a concept of a value that admits of comparative degrees. When using this concept, Kant never says that one action has *more* or *greater* moral worth than another. He says instead only that actions done from duty *do have* (true, inner, authentic) moral worth, while dutiful actions that are not done from duty *do not* have it, or have *no* moral worth. The error to be avoided here is not only the extremely common one of thinking that Kant regards dutiful actions not done from duty as morally *worthless* (or even immoral)—which would be self-contradictory, as long as we assume that *conformity to duty* is itself a kind of moral value. Also to be avoided is the far more subtle error committed by Marcia Baron when she attributes to Kant the view that “acting from inclination . . . is in some important way not *as good*—not *as morally worthy*—as acting from duty” (my emphasis) (Baron, 2002, p. 94). Kant’s claim that actions done from a laudable inclination, such as sympathy or love of honor, do not have moral worth is *not* the claim that they have *less* moral worth than the like actions done from duty. It is rather the claim that they *entirely lack* that special *kind* of worth that those actions do have. They may, for all that, be highly valued from the moral point of view; Kant says as much when he asserts that such actions are deserving of praise and encouragement.

1.3 Moral Worth is Different from Merit

Kant perhaps makes a mistake in not calling attention to the peculiarity of this concept of “authentic moral worth.” Instead, he focuses on the way it is of theoretical interest to him—namely, its role in deriving the first formula of the moral law. A second feature of the concept of moral worth that Kant doesn’t highlight is that it is not a concept that moral agents themselves have reason to employ in considering their actions. This too, I think, is responsible for many misunderstandings of Kant’s discussion at the beginning of the *Groundwork*, because readers often

assume, without thinking, that the concept of moral worth is something more familiar to us than it is. Readers familiar with ethical theories oriented to maximizing the good often even suppose that moral worth must be something Kant thinks we should pursue or even try to maximize in actions. One consequence of this mistake is that they tacitly identify moral worth in his sense with some quite different concept belonging to the family of concepts that give expression to distinctively moral value. The concept most often unreflectively chosen for this role, I believe, is the one Kant treats under the heading of moral “merit” (*meritum*, *Verdienst*) (See Johnson, 1996, pp. 313–37). We are tempted to think, for example, that Kant’s point in the comparison of the friend of humanity who is beneficent from inclination with the same man, when sorrow has overshadowed his inclinations and he must act from duty, is that the man deserves greater praise in the second case than the first, and his action has greater merit. Kant may think this about that example, but that is not his point in his discussion, and it is definitely not what he means by saying that his act has *authentic moral worth* only in the second case. An examination of Kant’s conception of moral merit should help us to see this.

Kant’s treatment of merit begins with the use of such a notion in the law, where merit is a concept of desert, namely the polar opposite of the concept of deserving punishment. Before a legal forum, an action with a certain kind of negative moral value (= –a) deserves punishment (*poena*), while the action with the opposite positive moral value (= +a) deserves a reward (*praemium*) (MS 6:227–8). In this context, Kant’s theory of merit belongs to his theory of the imputability of actions and consequences, where the evil consequences of an action in violation of duty are to be imputed to the agent, while its good consequences (if there are any) are not; the consequences of an action that is merely owed are not to be imputed (whether they are good or bad); and of the consequences of a meritorious action, only the good ones, not the bad ones, are imputable to the agent (MS 6:223, 227–8). In the moral or ethical sphere, an action that violates a duty deserves blame or disapprobation, whereas an action possessing moral merit presumably deserves praise or honor (cf. MS 6:384; R 6:22n). In the case of morally good actions that benefit a given person, for instance, Kant says that a meritorious action places the beneficiary under an obligation of gratitude to the agent (MS 6:448). But the notion of merit plays an important role in Kant’s theory of ethical duty generally, on account of its relation to his important distinction between wide or imperfect duty and narrow or owed duty, since the former are also regarded as *meritorious* duties (MS 6:390–1).

Unlike moral worth, merit does admit of degrees. Kant identifies several different factors as tending to make an action more meritorious. First, an action has greater merit to the extent that it does more than can be required (or in the legal case, compelled) (MS 6:227). Second, an act has greater merit the more difficult it

is, and especially if its performance requires the overcoming of subjective obstacles (the resistance of recalcitrant inclinations, non-moral reasons, or other motivational hindrances) (MS 6:228). Third, and apparently related to this second point, an action has greater merit if it is done at greater self-sacrifice to the agent. Finally, perhaps based on the first point, a morally praiseworthy action acquires greater merit the less it is demanded by strict duty, and the greater the readiness of the agent to perform it even so. In this respect, duty and merit are *contrasting* concepts: the more an act is required by duty, the less merit, and vice-versa: "If, . . . with no obligating grounds, [one person] acts well toward the other, simply because he wanted to, then the duty is smaller and more merit is ascribed to the act" (VE 27:568; cf. MS 6:228). As this last remark illustrates, and as Johnson perceptively observes in his article, Kant's distinction between wide and narrow duties should be regarded as a matter of degree or gradation: some wide duties are wider than others, providing more scope for meritorious conduct (cf. MS 6:390) (Johnson, 1996, p. 320).

This last point provides us with an obvious difference between the concept of merit and the concept of the moral worth of actions that Kant is using in the early pages of the *Groundwork*. On several different dimensions, merit is a matter of *degree*, whereas moral worth (as I have already argued) designates only a special *kind* of moral value. Thus, I suspect Baron of sliding from the concept of moral worth into the more familiar concept of merit when she ascribes to Kant the view that dutiful actions done from duty have *greater* moral worth than those done from inclination. As a matter of fact, however—and this is another important point of difference—the merit attaching to an action is for Kant to a considerable extent independent of the question whether it is done from duty.

There are, of course, at least two reasons for thinking that an action done from duty will *also* be to some degree meritorious. First, Kant thinks there is a duty to act in conformity with duty from duty, and this is a wide or meritorious duty (MS 6:391). In other words, someone who does what duty demands, regardless of the motive, has thereby done all that morality requires; but if, in addition, the agent does it from duty, this goes beyond what is required, and is therefore meritorious. This point, made long ago by Paul Dietrichson (1962), is necessarily missed by those who fall into a crude misreading of Kant that ascribes to him the view that there is a strict duty to act from duty, so that any act not done from duty is wrong or blameworthy. Second, as we shall see later, Kant's conception of acting from duty, as it is employed early in the *Groundwork*, applies only to cases where there is some impulse, inclination, or reason *not* to act in conformity with duty, or at least where there is no non-moral reason to act according to duty. This implies

that acting from duty will probably be subjectively difficult, requiring the agent to overcome subjective obstacles to the performance of duty, and this struggle of self-overcoming, as we have seen, is another source of merit.

Nevertheless, it is important to appreciate that actions can very well be meritorious, even highly meritorious, even if they are not done from duty. An action for Kant is more meritorious the less strict the duty requiring it, and the greater the subjective difficulty and the self-sacrifice involved in its performance: this is the case *whether or not the action is done from duty*. Love of humanity itself, Kant says, is always meritorious, because the duty we have to it is wide or imperfect (VE 27:669). So if readers of the opening pages of the *Groundwork* take away from it the impression that Kant thinks there is less merit in beneficence performed from love than in the performance of a strict duty from the motive of duty, then they get his views *exactly wrong*. These views entail that a beneficent agent who helps another out of sympathy, love, or inclination, but at great self-sacrifice, performs an action that is far more meritorious than one who helps from duty but at very little cost to himself. Either is certainly more meritorious than the act of an agent who acts from duty but gives the person only what is strictly owed. If the latter action has any merit at all, it could be only because of the degree of subjective difficulty or the self-sacrifice involved in conforming to the strict duty.

Kant says that “a man who himself has to overcome hardships, yet supports a person in need, has merit beyond that of a rich man, or one who pays his debts on time; for the well-doing is much less of a duty than the legal obligation arising from the right of another, even when it is difficult for the debtor” (VE 27:567–8). This is equally true, perhaps even true to a greater extent, when the poor but generous and self-sacrificing person acts from love or some other benevolent inclination than when she acts from a sense of duty. For the greater the strictness of the duty, the less merit there is in its performance. It is nevertheless true that, in terms of the concept of moral worth that is of crucial importance at the beginning of the *Groundwork*, the highly meritorious beneficent action of the self-sacrificing agent who acts from generous love has *no* authentic moral worth, while the action of the agent who pays a minor debt from duty at no sacrifice to himself *does* have true moral worth. Moreover, I think we (and Kant as well) would regard an act done from inclination and—going far beyond what the agent owes anyone—benefits many people at great self-sacrifice, as having far greater value from the moral standpoint than the act of the agent who, acting from duty, merely pays a small debt at trivial cost to himself. But in terms of the concept of moral worth that figures at the beginning of the *Groundwork*, the *kind* of value pertaining to the latter’s action is authentically moral, while the highly meritorious action of

the former agent is not. I suspect that many readers of the *Groundwork* think that because Kant denies moral worth to the former action, but accords it to the latter, he disagrees with the judgment that morality values more the meritorious act done from inclination than the act of no merit done from duty. I hope they can now see that this is a serious misunderstanding of his position. That is why it is an error to think that Kant's point in the example of the sorrowful man who acts from duty is that his beneficence has greater merit than it did when he acted from inclination. I hope they can now see that whether or not Kant would have made this judgment about relative merit, it is not what he is trying to say in this discussion.

1.4 Moral Praise is Distinct from Esteem

Kant says that a dutiful action motivated by an inclination such as sympathy or love of honor “deserves praise and encouragement, but not esteem” (G 4:398). Kant regards *esteem* as an attitude distinct from *praise* that is appropriate to the special *kind* of value possessed by actions with moral worth. Esteem is directed towards a *higher* (because more properly moral) kind of value, but not towards a *greater* value. Esteem, in other words, is *not* a *greater degree* of praise or approbation. It is an attitude of moral value that is different in kind from any attitude appropriate to a meritorious action that deserves praise. Thus Kant says of the action with moral worth that its worth is “the highest without any comparison.” Moral worth is a value *higher in kind* than any value that a dutiful action could have if it is not done from duty. But it is *not* greater on some scale of praise or approval that actions with authentic moral worth *share* with dutiful actions which are not done from duty. The latter actions, however praiseworthy and meritorious they might be, lack authentic moral worth.

Hence it is also not true that in Kant's view an action with moral worth is always to be regarded as *more valuable*, from the moral point of view, than an action lacking moral worth. This would be true only if all acts having moral worth are, from the moral standpoint, to be ordered lexically ahead of all acts lacking moral worth. But Kant never says anything to this effect, and on reflection, in light of the examples given, such a claim would seem extremely implausible. It would also directly contradict Kant's theory of merit. As noted earlier, we do not think that the performance of a small but strict duty, from duty, is more meritorious than a self-sacrificing action of great benefit to others not strictly required but performed from a generous inclination. The chief error to avoid here is that of interpreting Kant's statements in terms of a shallow, simplistic, one-dimensional conception of moral value, so to speak, instead of recognizing the complexity of the set of concepts needed to account for the varied kinds of moral value that there are.

1.5 “From Duty”: What Kant Does and Does not Mean by This Phrase

I believe we can fully appreciate this complexity only if we turn to the other main object of misunderstanding in Kant’s discussion: namely, his conception of what it is to perform a dutiful act “from duty.” This conception too, is in my opinion very often badly misunderstood. The most common source of misunderstanding is the natural but mistaken supposition that when Kant talks about acting from duty, he must have in mind a scenario in which the agent’s action is (as people say) “motivationally overdetermined”—in other words, where the agent has at least two incentives for performing the same action, and the question (one of moral psychology) is which one of them “really motivates him.” A look at Kant’s examples, and careful attention to what he says about them, should suffice to discredit this supposition.

Kant’s first example is that of the merchant who holds to the policy of not overcharging inexperienced customers and charging all his customers the same honest price. Presumably the merchant has a moral incentive for this policy of honesty, but Kant says that the merchant realizes it is also the most prudent policy, and concludes *directly, from this alone* that the merchant’s honest action was done not from duty but with a self-serving aim. If Kant were interested in the psychological question of the agent’s “real motive” in cases of “motivational overdetermination,” he might have to consider it at least possible that the merchant is acting from duty, even though he has a prudential motive as well, and he should have mentioned this possibility. But he does not. Instead, he infers immediately—merely from the fact that the merchant *has* a potent self-interested reason for his policy of honesty—that his action is *not* done from duty.

It is similar with all the other examples Kant takes up, where the agent is supposed to have not a prudential reason but an immediate inclination to a dutiful action. (These examples are those of preserving one’s life, helping others in need, and fulfilling the indirect duty to provide for one’s own happiness.) In each case, if the question were about the agent’s “real motive,” Kant would need to consider two possibilities: first, that the agent is doing the dutiful action from duty and second, that the agent is doing it from the immediate inclination. But in all these cases, Kant infers *directly, merely from the presence of the inclination*, and without any further speculation or stipulation about the agent’s motivational dispositions or priorities, that the dutiful action is *not* done from duty. *In no case does Kant consider someone having both duty and some other incentive to perform a dutiful action whose action has moral worth because they act from duty rather than from some other motive.*

I submit that this is because *he is simply not using the term “from duty” to refer to the agent’s “real motive”* in cases where more than one motive for the same dutiful action is possible. Instead, he is using the term “acting from duty” in such a way that it is possible for an agent to “act from duty” at all *only* when there is no other incentive for the dutiful action except the motive of duty.

We can find further confirmation for this if we pay close attention to the thesis Kant later states as his “third proposition”: “Duty is the necessity of an action from respect for law.” Admittedly, the meaning of this proposition is not transparent. Many think it must mean only that our duties are actions that we *must do* because the moral law commands them. This certainly sounds like something many people might take to be true, perhaps even self-evident. In some contexts, it might even be what Kant would be asserting using these words. But on closer consideration it cannot possibly be what Kant means here. Wide duties, such as the duty of beneficence Kant has just been discussing, are not morally necessary or must-do actions. So if that is to illustrate Kant’s third proposition, it cannot mean that duties are must-do actions.

In the context of the early pages of the *Groundwork*, I submit that Kant’s third proposition makes most sense if we take it to be a claim directly about the topic he has just been discussing, namely a claim about what it means to *act from duty*. Kant is saying that *to act from duty* is to necessitate (compel or constrain) oneself to do something out of respect for the moral law. This could apply equally to strict or owed duties (actions that are morally necessary or must-do) and to wide or meritorious duties (which are duties, but not morally necessary or must-do actions). For it is possible to *constrain yourself* to do a meritorious action out of respect for the moral law. This further agrees with what Kant says about duty at the beginning of the Doctrine of Virtue: “The very concept of duty is already the concept of a necessitation (constraint) of free choice through the law... [which] may be a *self-constraint*” (MS 6:379).

It is sometimes suggested that Kant declines to consider “motivationally overdetermined” actions having moral worth only because he thinks moral worth is more *conspicuous* in cases where there are no co-operating incentives. This suggestion, however, receives no support from the texts, and is even rebuffed by them. Kant does say that actions done from duty are cases in which *the good will* “shines forth more brightly” (G 4:397), but he does *not* say that actions done from duty without co-operating motives are cases where *moral worth* is more conspicuous. If this were what he meant to say, then he should have said that actions done from duty with no co-operating non-moral incentive are actions that *exhibit* their moral worth more conspicuously. But he does not say this. Instead, what he says is that the actions *have* moral worth at all, and their maxims moral content, *only* when all such incentives have been removed (G 4:398–9).

To act from duty, then, means simply *to act with moral self-constraint*—making yourself do an action because of that action’s relation to the moral law (which the law says is either a narrow or a wide duty), and out of respect for the law. But self-constraint is possible only when there is no other incentive sufficient to motivate the action. If you have some other incentive to do the action, such as an immediate inclination, then making yourself do it from duty is no longer necessary; indeed, it is not even possible. You *can’t* make yourself do something, in other words, unless you *need* to make yourself do it if it is going to be done at all. This explains why the mere presence of a self-serving motive (in the case of the merchant) and immediate inclination (in the case of Kant’s three other examples) is enough to guarantee that the dutiful action is not being done “from duty”—*in the sense of that phrase Kant means* at the beginning of the *Groundwork*. In each of Kant’s examples of action done from duty, he imagines the action as done from duty simply by imagining the agent doing it after the removal of the immediate inclination to do it—for this leaves as the only possibility that if the action is to be done at all, it must be done because the agent constrains himself to do it out of respect for the moral law.

The most interesting claim Kant is making early in the *Groundwork* is therefore this: an action has moral worth—the *kind of value* that lies right at the heart of morality—only when it is an action the agent has had to make himself perform out of respect for the moral law. *Morality*, in other words, is essentially about rational self-constraint from respect for the moral law. Although there are many other valuable things that morality may care about, at its very center morality is about nothing other than rational self-constraint from respect for the moral law. I think it is a shame that most readers of the *Groundwork* are so distracted by irrelevancies and misunderstandings that few of them even get as far as entertaining the basic idea Kant is trying to put before them in the early pages of the *Groundwork*. When they show any awareness of this claim at all, they often do so only by misunderstanding it and confusing it with different claims that are easier to reject than it is.

1.6 How *Kantians* Usually Go Wrong Here

Even Kantians tend to misunderstand this discussion, because they try to see it as illustrating points of Kantian moral psychology that it is not meant to illustrate. It is an important Kantian doctrine, for instance, that in doing our duty, we should give preference to moral motives over other motives, and that we should encourage in ourselves the rational habit of doing this: according to Kant, that rational habit is the main thing that virtue or moral strength consists in. Kant does say at G 4:398–9 that the sorrowful man who is beneficent from duty exhibits goodness

of character (in other words, virtue). But in this discussion, *Kant is not interested in virtue*: i.e. the *habitual* strength of the motive of duty in the struggle with other motives, nor is he interested in the priority we ought to give to moral motives over non-moral ones. He is interested instead in the special *kind* of goodness (properly or authentically moral goodness) exhibited by a dutiful agent who needs to exercise self-constraint, in contrast to the agent—equally dutiful as far as maxims and conduct are concerned—who does not need to exercise such constraint.

When I point out that the honest merchant or the beneficent sympathetic friend of humanity have (for all we are told in this discussion) a will just as good as that of the agent who acts from duty, a common reaction from Kantians is that I am treating the motive of duty merely as a “default” or “back-up” motive—which, of course, would be profoundly opposed to what Kant thinks it should be. Actually, I have never claimed any such thing or even suggested it. I think they say that because they cannot get it into their heads that in this particular discussion, *Kant is simply not interested in expounding his views about the nature of virtue, or the rational priority of moral over non-moral motives*. Consequently, the contrast between the prudent merchant or sympathetic friend of humanity and those who act from duty is simply not intended to illustrate or invoke those doctrines, or to raise the issues those doctrines address.

1.7 The Duty to Act from Duty

These issues come up mainly in connection with Kant’s claim that “as to what is to be morally good, it is not enough that it *conform* to the moral law, but it must also happen *for the sake of this law*” (G 4:390) and his claim that we have a duty to act from duty (MS6: 383, 393, 394–5). It may seem self-evident that these claims must be closely related to—virtually the same in meaning as—his claim that actions have moral worth only if done from duty. Kant himself even seems to associate the two in connection with his distinction between the *legality* of actions (their conformity to the moral law) and their *morality* (that they have been done “for the sake of the law alone”) (KpV 5:81). Yet as soon as we investigate further what Kant takes the duty to act from duty to consist in, the association becomes extremely problematic.

The duty to act from duty is presented by Kant as a *duty of virtue*, an aspect of the duty to cultivate one’s moral perfection (MS 6:393). But sometimes it is also described as a duty pertaining to “what is formal in the moral determination of the will” or “what is formal in maxims” (MS 6:383, 395). Let’s first consider the duty as a duty of virtue. When it is considered this way, it is not a duty to be fulfilled through performing dutiful actions in a particular way—such as by choosing to be incited

to perform that particular action through one possible motive rather than another. Of the duty to act from duty, Kant says:

Hence this duty too—the duty of assessing the worth of one's actions not by their legality alone but also by their morality (one's disposition)—is only of *wide* obligation. The law does not prescribe this inner action of the mind but only the maxim of the action, to strive with all one's might that the thought of duty for its own sake is the sufficient incentive of every action conforming to duty (MS 6:393).

The duty to act from duty is described here in two ways: first, it is a duty to *assess* one's actions in a certain way, and second, it is a duty to *strive* to make the thought of duty the *sufficient* incentive of every dutiful action. One has a duty to assess one's own actions not only by their dutifulness but also by one's disposition. But Kant never says that one has a duty to assess one's actions according to their *moral worth* (in the sense of that phrase used in the *Groundwork*). It is only we (as representatives of common rational moral cognition, considering these examples) who even notice the difference between actions with authentic moral worth and dutiful actions that do not have it. One might also assess one's actions according to one's disposition, rather than merely according to their external conformity with the moral law, even if in a given case there is no fact of the matter whether one acted from one incentive or another in performing the action. If one has striven for the right disposition, and the strength of character to give duty priority over competing motives in cases of conflict, then that by itself constitutes meritorious satisfaction of the duty to act from the motive of duty. It guarantees that duty is not merely a “back-up” motive, even in cases where one acts dutifully from inclination rather than from duty, because inclination makes it unnecessary to constrain oneself to act according to duty.

Notice too that one can strive (even with all one's might) to make the thought of duty a sufficient incentive, even if it never is in fact a sufficient incentive. This duty to do one's duty from duty unlike the moral worth that belongs to actions done from duty, is also not about particular dutiful actions, but concerns an end that pertains to one's general motivational state. One can comply with the duty to act from duty even if there is no fact of the matter in each particular case of dutiful action about whether one acts from the incentive of duty or from some other non-moral incentive. Further, in this passage Kant explicitly *denies* that the duty to act from duty is a duty to perform an “inner action of the mind”—such as the act of choosing one incentive over another on which to perform some action in conformity with duty. That is precisely why the duty to act from duty is a wide, rather than a narrow, duty. For wide duties are duties to adopt an end (in this case, the end of making duty a sufficient incentive) and narrow duties are duties to perform

or omit a particular action (such as the “inner action” of choosing to act for one incentive rather than another when one has two different incentives for the same dutiful “outer” action).

Just prior to the quoted passage, Kant has even given an *argument* for the denial that the duty to act from duty is a narrow duty to perform such an inner action:

For a human being cannot see into the depths of his own heart so as to be quite certain, even *in a single action*, of the purity of his moral intention and the sincerity of his disposition, even when he has no doubt about the legality of the action . . . In the case of any deed it remains hidden from the agent how much moral content there has been in his disposition (MS 6:392–3).

Kant’s reasoning here, as I understand it, is the following: if there were a strict duty to choose the incentive of duty over other incentives in the case of each dutiful action, then there would have to be a discernible fact of the matter whether one had done this. But the only fact here is not a fact about such an inner act, but a fact only about one’s character or *disposition*. This is not a fact about a single action, and it is even a fact impossible for a human being to determine with certainty in a single action. Therefore, the duty to act from duty cannot consist in a narrow or strict duty to perform an inner action of choosing the incentive of duty over other incentives. (There is no such strict duty.) Instead, this duty must consist only in a wide duty to perfect one’s virtue (the strength of one’s good maxims) so that, if one’s striving were perfectly successful, the incentive of duty would be sufficient for the performance of duty even in cases where there are no co-operating incentives but in fact contrary incentives with which to contend. It is meritorious to engage in this striving to perfect oneself, even if the striving falls short of its goal. Thus it cannot be contrary to duty to need moral incentives to do one’s duty—as long as one actually does it—and *a fortiori* there can be no strict duty to be motivated only by duty in performing one’s duty in a particular case. To understand the duty to act from duty in the latter way, therefore, is to misunderstand it. It is to take this duty in a sense that would make it not only excessively strict, but even impossible to fulfill (perhaps even in a way that makes the concept of its fulfillment nonsensical).

Another way to put this point would be that when I say something like: “She not only did the right thing (this particular dutiful action) but she also did it for the right reason [i.e. from the right incentive or motive],” my remark must always be about *her* (her character and disposition) and cannot be only about *this particular action* (or its imagined motivational psychology, taken entirely in isolation from the kind of person she is). To be sure, a vicious person can sometimes do a virtuous action, by acting (as we say) “out of character.” But I do not think it makes sense to claim that a person whose character and moral disposition are generally bad

performed this particular dutiful action from the right motive (e.g. from duty), even when she had a sufficient motive other than duty, and even though her character disposes her to perform actions of that kind from this less creditable motive.

Suppose M is a merchant who usually gives correct change to children from the motive of business prudence, but (*unlike* Kant's merchant at G 4:397) would (also from self-interest) cheat them if she thought she could get away with it. Suppose, with this understanding of M's character as the background, someone said: "M just gave little Lucy the correct change, and she knew Lucy's mother was watching her, so she knew it would have been imprudent to cheat Lucy, but this time she did it from duty, and not from any selfish motive." It is not merely that I do not know on what evidence the person might make this claim about this particular action, I doubt that the person's claim even has truth conditions. I might perhaps make some sense of the remark if it was part of a story that M had undertaken to reform her bad character, and the person is trying to tell me that this action represents a stage in that process. But I don't think the remark makes any sense if the claim is that M, irrespective of her basically selfish and dishonest character, made a wholly isolated choice of this dutiful action not from selfish motives—though she certainly had ample selfish reasons for doing it—but instead performed this one honest action from duty because (again, acting wholly out of character) she chose the motive of duty over the motive of business prudence on this one isolated occasion. I think there would then be no truth conditions for that claim.

There are accounts of Kant's moral psychology that apparently do try both to understand the duty to act from duty as the duty to choose the motive from which one does the dutiful action, and also regard Kant's claim that only actions done from duty have moral worth as a claim about actions that fulfill the duty to act from duty understood in this way. One such account is given by Henry Allison, first in his book *Kant's Theory of Freedom* (Allison, 1990), and then more recently in his *Commentary on the Groundwork* (Allison, 2011, pp. 114–20).

This reading is based on what Allison calls Kant's "incorporation thesis." Some of what Allison claims here has firm textual support and is beyond dispute, so I should begin by acknowledging that part of what he says. The incorporation thesis, as stated in the *Religion*, holds that no incentive can determine the human power of choice except insofar as the agent has incorporated the incentive into his maxim (R 6:24). This thesis is a rejection of a view of action commonly taken by philosophers downwind from Hume, who hold that desires can directly cause actions. Kant's contrasting view is that the determination of the will by incentives is compatible with freedom only if we suppose that an incentive determines the will through the choice of a maxim in which the incentive has been incorporated. Allison argues, and I think he is correct, that this is evidently Kant's view not only

in the *Religion* (where the incorporation thesis is first explicitly stated) and afterwards, but also that it was implicit in Kant's earlier ethical writings as well. It is further Kant's view in the *Religion* that in order to account for evil in the human will, we must suppose that human agents have a propensity to volition according to the maxim of preferring incentives of inclination or self-love over the (rationally superior) incentive of morality, so that at least in the case of this maxim, there can be the incorporation of more than one incentive, and also a determinate order in which the maxim prefers one incentive to another.

From these points, however, Allison seems to infer a peculiarly strong version of the incorporation thesis, which does not seem to me to follow from the incorporation thesis as it is stated in the *Religion*, and is even incompatible with Kant's moral psychology more generally. Allison holds, namely, that the incorporation thesis entails that for *every* maxim adopted by an agent, and even for every action, there must be some explicit choice of moral over non-moral incentives, or vice-versa. In other words, Allison thinks it is Kant's doctrine not only that it is a necessary condition of acting on an incentive that it is incorporated in a maxim, but also, conversely, that every maxim, and every action on a maxim, has to involve such an act of incorporation. Thus every maxim adopted by an agent must include not only the choice of a way of acting, but also the choice of an incentive—hence a choice to prefer *this* incentive over other incentives incorporated in the maxim along with the preferred incentive.

On Allison's view, it is true not only that the human propensity to evil, and the moral disposition to combat it, involve maxims involving a determinate order of moral and non-moral incentives, but also that every maxim, and every action involves such a choice. It would be impossible, on Allison's view, for someone to adopt the same maxim either from one incentive or another, or to act on the same maxim sometimes from one incentive and sometimes from another. For the incentive, and the priority among incentives, would be built into the maxim, and a different incentive or priority among incentives would mean a wholly different maxim.

Thus Allison thinks there are two possible variants on the case of the honest merchant discussed by Kant at G 4:397. The merchant can either act honestly from prudent self-interest (rather than from duty) or act honestly from duty (rather than from self-interest). And one of these variants (rather than the other) must hold in every case where the merchant gives change. It can't be merely that the merchant acts honestly, having both a moral and a non-moral incentive for so doing. Allison rejects Marcia Baron's suggestion that there could be no such thing as refusing to act on an incentive, choosing *not* to act on it, while acting in a way that the incentive gives you an incentive to act (Allison, 2011, pp. 115 n. 49, 116–20). On the contrary, the merchant must either choose honest dealing along with the incentive of

duty, or choose honest dealing while preferring the incentive of self-interest over that of duty.

On Allison's reading, then, it seems evident that along with every duty, whether wide or narrow, to act in a certain way (to deal honestly, to preserve one's life, to act beneficently, etc.), there is also a *strict* duty so to act from duty—that is, to incorporate the incentive of duty into one's maxim, and to give it priority to any other incentive one might have for fulfilling the same duty. This duty to act from duty must be a narrow or strict duty, for if one does not incorporate the motive of duty in one's maxim, and give it priority over all other incentives one might have for being honest, beneficent, etc., then one has made an *evil* choice, a choice in direct violation of the moral law.

Allison's account does not fit what Kant says about the various examples he considers at G 4:397–8. If Allison were correct, Kant should have explicitly raised the question whether the merchant acts from duty or from prudence, and perhaps allowed that the merchant might be acting from duty, instead of concluding, as he does, straightaway, from the fact that he has prudential grounds for the maxim of honesty, that he is acting not from duty but with a self-interested aim. He should also have asked the same question about the other examples (preserving one's life, acting beneficently, etc.) Kant should have pointed out that if (as he says) most of us most of the time do preserve our lives from immediate inclination, then we act wrongly and blamably. And Kant should *not* have said, as he does, of the friend of humanity who helps others from sympathetic inclination that his beneficent action is "dutiful and amiable, deserving praise and encouragement." On the contrary, on Allison's interpretation he should have condemned this man, not for his beneficence, but for his choice of incentives, and neither praised nor encouraged his conduct, but instead should have warned us against that choice of incentives as directly contrary to duty.

Allison's excessively strong interpretation of the incorporation thesis requires him to contradict Kant's claim that the duty to act from duty is a wide duty, and not a duty to perform an inner action. But Allison might still claim that what he says applies the duty to act from duty considered not as a duty of virtue, but as a duty pertaining to what is "formal in the determination of the will," or "formal in maxims"—which, he might say, refers to the choice between moral and non-moral incentives which occur in every maxim according to his understanding of Kant's incorporation thesis. This, however, would still conflict with Kant's emphatic assertion that the duty to act from duty is a wide duty rather than a strict duty, and his explicit denial that it is a duty to perform an "inner action of the mind."

There is no sign, moreover, that Kant considers the duty to act from duty as what is "formal in the determination of the will" to be a *different duty* from the wide or

imperfect duty to promote our moral perfection by striving to make the thought of duty a sufficient incentive. Rather, the latter duty is simply a duty to promote a virtuous disposition in ourselves, and it is this disposition that manifests itself in the “formal determination of the will” or what is “formal in maxims.” Kant introduces “what is formal in the determination of the will” by saying that it “corresponds to the concept of virtue,” associating this duty with “what it is virtuous to do” (MS 6:394). In considering the question whether virtue is one or many, Kant contrasts the “virtuous disposition,” which can be considered a single unifying virtue, with the plurality of more particular virtues that might correspond to different virtuous ends or duties of virtue (MS 6:383).

If the duty to act from duty is really a wide duty to act so as to further a certain character or disposition in ourselves, rather than a duty to perform some special inner (motivational) act in addition to the external act that conforms to duty, then we can see that there is no supposed “regress” to which Kant is subject in holding that we have a duty to act from duty. (For the bogus claim that Kant is vulnerable to such a regress, see Ross, 1930; Allison, 1990, p. 178; and Pippin (in Guyer (ed.)), 2006, p. 442 n.9.) Even when considered as what it is “formal in the determination of the will,” the duty to act from duty is not thought of as the performance of an inner action of the mind or the choice of one incentive over another in performing a specific dutiful action. It is instead a general feature of the agent—namely, *virtue* (as strength of good maxims) or the *virtuous disposition* that may manifest itself in particular actions as the “formal determination of the will.” To fulfill the duty to act from duty is *never* to perform a specific inner act of preferring one incentive over another with reference to a particular dutiful action. It is therefore *never* something that could pertain to a specific dutiful act, which could constitute the “moral worth” of that act. Therefore, there has to be a gap or mismatch—not an inconsistency, only a difference in subject-matter—between the issues discussed at G 4:387–98 and the issues involved in the duty to act from duty. The former has to do with whether a *particular action* has or lacks moral worth. The duty to act from duty, however, is never a duty that concerns a particular action. It concerns instead the cultivation of the agent’s disposition, character, or virtue, or the assessment of our actions in terms of this disposition. And the duty to act from duty is always a wide or meritorious duty, never a strict duty requiring an act of choosing one incentive over another.

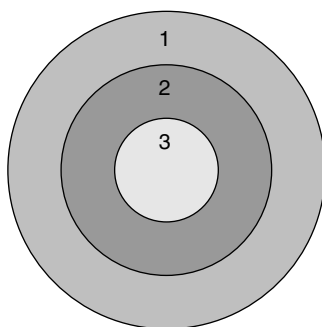
I find no mention anywhere in Kant’s writings of acts of choosing duty over other motives regarding the same dutiful action; I think his moral psychology is at least consistent with a refusal to admit that such acts ever occur. The evil maxim of inverting the moral order of incentives, discussed in the *Religion*, is not a specific inner volitional act accompanying our external actions, but rather an evil

propensity that is inferred from evil acts as the ultimate ground in free choice that they presuppose (R 6:20); and the contrary maxim that combats this propensity also does not pertain to particular acts, but rather belongs to an underlying pure *disposition* that the agent manifests in a changed course of life (R 6:46–7). It cannot be the case, therefore—however obvious it may appear to be—that “acting from duty,” in the sense that gives particular actions moral worth at G 4: 397–8, is just an instance of the fulfillment of the duty to act from duty. The moral worth of actions done from duty and the duty to act from duty are quite different topics in Kantian moral psychology; their confusion is the source of much mischief in reading the opening pages of the *Groundwork*.

Thus, when I say that in the terms of the early pages of the *Groundwork* the fact that a person can perform dutiful actions—even perform them with a good will—without needing to constrain themselves from the motive of duty, this does not imply that I think in those pages the motive of duty is being regarded in general only as a “default” or “back-up” motive for moral actions. The point is instead to distinguish—even perhaps among agents who have an equally good will or an equally strong moral character, or an equal commitment to the priority of moral over non-moral motives—between cases in which their dutiful conduct exhibits what is central and most proper to morality and cases in which it does not. This is not a distinction that involves the nature of virtue or the priority of moral to non-moral motives. What I am saying is that to interpret Kant’s claims about “moral worth” as engaging those issues, or as stating familiar Kantian positions on them, is to *misinterpret* those claims. When we read the examples as illustrating Kantian doctrines, even important ones, that they are not intended to involve, we misunderstand them just as surely as we would if we read them by importing a moral psychology alien and hostile to Kant’s. So when I make this point concerning what this discussion is about, I am not *denying* the Kantian doctrines about the correct priority of motives. I am saying only that these doctrines are not the point of the discussion here—and also that the doctrines are misunderstood if taken to require us to make a choice between motives in particular cases where more than one incentive for the same action is present. The aim of the discussion in the *Groundwork* is rather to bring out what morality itself is about, and to delineate, if I may put it this way, Kant’s view of the structure of the things morality cares about.

1.8 The Structure of the Values Morality Cares About

Let me summarize the result of what I have been saying so far by suggesting the following image: morality may be viewed as a circle, or a series of concentric circles.



1. Non-moral values which morality nevertheless cares about, such as human health, flourishing, well-being and happiness, and human talents, qualities of mind and temperament that promote human ends. These belong to well-being (*Wohl*), the worth of someone's state or condition (*Zustand*) (KpV 5:60), which morality values, either as a duty of virtue: an end which is our duty—the happiness of others (MS 6:385–8) or because a moral agent's conduct has made the agent worthy of happiness (G 4:393).

2. Values displaying good will—conformity to duty, moral merit, the priority of moral to non-moral motives in character, virtue, and particular virtues, such as beneficence, truthfulness, respect for human dignity. These are properly called good (*Gut*) because they belong to the goodness of actions, or the way of acting, or the acting person as a good human being (KpV 5:60). These constitute the worth of a person. All of them in one way or another involve the unlimited goodness of the good will. But they still do not have that very special kind of worth that belongs centrally and authentically to morality. (After all, God has a good will too, but God is not a moral being. God's actions lack moral worth, because God cannot act from duty.) What is most inwardly essential to morality, in Kant's view, is only for imperfect beings (human beings), who must constrain themselves through moral necessitation.

3. Authentic moral worth—acting from duty. This is absolutely central to moral value, the only thing truly having the special kind of value that belongs to morality. This worth belongs to the good will, but strictly speaking only in its true, inner or authentic moral form; that is, under conditions of adversity, “which, however, far from concealing the good will and making it unrecognizable, rather elevate it by contrast and let it shine forth all the more brightly” (G 4:397). It is from this absolutely central moral value that Kant hopes to derive the moral law in the First Section of the *Groundwork*.

At the center is the circle containing the value Kant considers most basic: (3) the *moral worth of actions done from duty*. Surrounding this central circle is (2), a circle containing a larger set of values displaying *goodness of will*—conformity to duty, moral merit, the priority of moral to non-moral motives in character, virtue, and particular virtues: beneficence, truthfulness, respect for human dignity, and so on. Then still farther out from the center lies a still larger circle (1), comprising *non-moral* values, which morality nevertheless cares about, such as human health, flourishing, well-being and happiness, and human talents, qualities of mind, and temperament that promote human ends, which are good from the standpoint of morality. These are good from a moral standpoint when they are combined in the right way with moral goods—for instance, when talents are placed in the service of morally good ends and when people enjoy the happiness of which their conduct has made them worthy.

It is also important to appreciate, however, that while this set of concentric circles describes the *structure* of those kinds of valuable things morality cares about, it does not tell us which instances of each value are, from the moral standpoint, to be given preference over which others. As I've already suggested, morality may value the merit of an action done from sympathy that benefits many people more than it does the moral worth of an action done from duty that discharges only a small obligation of justice (such as the payment of a minor debt). And if a moral agent is dedicated to a meritorious end belonging to the outermost circle—for instance, relieving the suffering of many people—then she will naturally care far more about this end than she does whether some of her actions taken towards it have moral worth because they are done from duty. Indeed, Kant's theory does not justify the agent's concern with this at all, unless the case is one where she will fail to act in conformity with duty unless she acts from duty, and then it is dutifulness itself, not action from duty or moral worth, that matters to the agent.

Discussions of the beginning of the *Groundwork*, as I have said, often treat moral worth as something agents have reason to want their actions to have. But as I have also said, I think this is a thoroughly misguided thought. In none of Kant's examples does he portray the agent who acts from duty as caring about the moral worth of actions. The interest of moral worth for Kant is a purely theoretical one: to appreciate the centrality of this value to morality helps us to understand morality better. It would do us no good as moral agents to fixate on this value as something we ought to try to exemplify, still less to maximize. A conscientious moral agent does have reason to want their actions to conform to duty, and hence also to be the kind of person who will still do her duty when they must constrain themselves on moral grounds to do it. In that sense, we should want our actions to have moral worth *when they must have it if we are to do the dutiful thing*. But no moral agent

ever has any reason to want to be in a situation where such self-constraint is necessary. On the contrary, we all have reason to want it to be easy and pleasant to do what we ought. In that sense, we have every reason to hope that we will seldom be in situations where our dutiful actions have moral worth, and no reason at all to want our actions to have moral worth if we can do our duty without our actions needing to have it. Moral worth, we might say, like the goodness without limitation of the good will, is primarily an “appraiser value,” not an “agent’s value”—not a value employed by moral agents in the course of deciding what to do.

To say this, I repeat, is not to treat duty as merely a “back-up” motive, nor to deny that we should try to do our duty from duty, rather than from other motives. It is only to say that the discussion of acting from duty and moral worth at G 4: 397–9 is *not about those issues*, and that to think that it is, is to miss the point that Kant is trying to convey to us in that discussion.

1.9 Moral Worth, the Good Will, and the Formula of Universal Law

The main theoretical purpose for Kant of the idea of moral worth is its connection with the derivation of the Formula of Universal Law. This connection is quite straightforward. As Kant argues at G 4:402, that formula tells us to restrict ourselves (constrain ourselves to adopt only) maxims that conform to the idea of universal law. So it follows in an obvious way from the conception of acting from duty, and hence from what has moral worth. The same idea also connects to the starting point of the *Groundwork*’s discussion, the idea of the good will (and more generally, to the thought that morality is about willing). The will is the capacity to govern our actions according to rational principles (see G 4:412). So if moral worth belongs centrally to actions involving self-constraint through respect for laws (objective practical principles), then morality obviously has to do with willing, and with *good* willing—or at least with a certain *kind* of good will, the kind that is good through moral self-constraint.

Not every case of the good will, however, even pertains to morality at all. The divine will, for instance, and more generally the holy will, is too perfect ever to require self-constraint, so the actions of God, or of any holy being, never have *moral* worth (*can* never have it). Morality, in other words, is not for beings with perfect wills: it is only for volitionally defective rational beings, for screw-ups like us, who often need to *make* ourselves do something (out of respect for rational principles) if we are going to act according to reason.

For exactly the same reason, it is not central to *morality* when we find the good thing easy or natural to do—as when we have a natural inclination to do

it. Theories like Hutcheson's and Hume's went wrong because they took the central cases of moral action to be the cases where doing the right thing or the good thing is natural (that is, easy or spontaneous). They therefore mistook moral virtue for a condition of innocence, or the possession of a fortunate temperament, a happy condition in which people do the right thing because they spontaneously like doing it. Kant's counter-claim agrees with them up to a point: Kant thinks true morality or virtue involves doing the right thing because you want to do it and because you value doing it for its own sake. That is precisely why Kant thinks it may be difficult for us, as it was for these moral sense theorists, to distinguish the action with authentic moral worth from a dutiful action for which we have an immediate inclination (G 4:397–8). We think that wanting to do a good action for its own sake is the same as having a natural inclination to do it. (Many misreadings of Kant also start from the supposition that when he uses the word "inclination" this is interchangeable with "desire"—forgetting that inclinations are *empirical* desires, while many of the desires that matter most—to means–ends and prudence, as well as morality—are *rational* desires.)

Dutiful actions done with ease and inclination do, in Kant's view, deserve praise and encouragement. What they do not deserve is the *esteem* we reserve for that special value that is central to *morality*. Here it is important to see that moral worth, which pertains only to acting from duty, is a value of narrower application than the goodness without limitation that pertains to the good will. Not only do God's actions display a good will without having moral worth: so do human actions that follow good maxims or principles (since will is action on maxims or principles, and a good will is action on good ones) but without the need for moral self-constraint. Not only God and the angels, but also the honest merchant who acts honestly from prudence rather than from duty, and the friend of humanity who is beneficent from sympathy rather than from duty, as well as most of us most of the time who preserve our lives and our happiness from immediate inclination, are examples of good will, but not examples of acting from duty or of moral worth.

God has a good will but is not a *moral* agent at all. (He is far above the human imperfection that *morality* presupposes.) Morality is also not involved, strictly speaking, in the good deeds of the sympathetic person or the one who does what is right when it is prudent or honorable. Morality comes on the scene only where innocence has been lost, where doing the right thing is no longer natural or easy, where doing the right thing goes against our natural grain. Morality properly applies only where nature, inclination, and temperament no longer assist reason and have ceased to provide any pleasant substitute for the painful business of making ourselves do what we ought to do.

That is what it really means to say that true (inner, authentic) moral worth belongs only to actions done from duty. Moral worth can show itself only under conditions of adversity—where we no longer preserve our lives because we have a natural instinct to do so, or help other people because we take pleasure in doing so, or even fulfill the indirect duty to promote our own happiness because we have a natural inclination to it. Morality appears when we find ourselves in the miserable situation where we come face to face with our own imperfections and must make ourselves do all these things because rational principles tell us to do so, either in order to avoid moral blame or to achieve moral merit.

Kant's view about what is central to morality is misconstrued when it is seen as involving a negative judgment about actions that conform to duty but do not require self-constraint in order to get done. People misread Kant, for instance, when they think that he is *criticizing* the honest merchant who deals fairly with even the most inexperienced customers, not through moral self-constraint but because he realizes that principled honesty is the most prudent and self-serving policy as well as the morally correct one (G 4:397). There is a tendency for Kant's readers to think that Kant is saying that this merchant must not really be an honest man at all: that his maxim is not to deal honestly but only to appear honest, and to behave honestly when he can't get away with being dishonest. Perhaps they think this because Kant says the merchant acts with a "self-serving aim." But for Kant there is nothing wrong with acting with such an aim, as long as your actions conform to duty—as the merchant's actions do, precisely because he sees it as prudent and self-interested to act honestly as duty requires.

A similar deafness to what Kant is saying leads people to think he has something against the sympathetic "friend of humanity" who "takes an inner gratification in spreading joy around him" and therefore helps other people from an immediate inclination to see them happy. When Kant says that such conduct is "dutiful and amiable," and "deserves praise and encouragement," some readers want to twist this into saying that it deserves only some sort of *non-moral* praise or *merely instrumental* encouragement. They think Kant regards it either as positively immoral or at least of no value morality cares about to do good to people unless, like the man Kant describes whose sympathies have been clouded over by his own sorrows, you do so gritting your teeth and acting from cold duty. They ignore Kant's claim that in being beneficent this man "tears himself out of deadly insensibility," which is obviously the diametrical opposite of remaining in that emotionless and unfeeling state. The motive of duty that most directly pertains to beneficence is surely the rational feeling Kant calls "love of humanity" (MS 6:400).

Or, alternatively, some attribute to the friend of humanity, as well as to the merchant, the evil tendency Kant calls "impurity," where one comes to depend on

motives of self-love or inclination to get one's duty done, which involves a weakening of one's character in cases where these helpful motives are not present and perhaps also to a distortion of one's perception of what one's duty is, so that one comes to see something as one's duty only when it is motivationally supported by inclination or self-interest (R 6:30). But to read such an invidious psychology into Kant's examples in the *Groundwork* is not only to see something that is simply not there, but it is also to take the examples in a way that distracts us from, or even blinds us to, the very point he is trying to make using them.

Kant's view about beneficence from sympathy and inclination is that, as far as it goes, this is a fine thing—it is often even conspicuously meritorious. There is nothing bad to be said about it, any more than about those who perform public-spirited good works because they desire the honor these works receive from their fellow citizens (G 4:398). In general, there is nothing negative to be said about anybody who does the right and dutiful thing because he naturally likes doing it, or even because it is in his self-interest to do it. Neither inclination nor self-interest “taint” good deeds. Kant's only point is that these actions do not exhibit that special kind of worth central to morality. This is not a *negative* point about prudent, natural, pleasant, or innocent goodness. It is rather a quite different point: about the nature of *morality*—namely, the point that at its heart *morality* is not about this kind of goodness, but about a goodness that is hard won, having to be fought for against our human imperfections.

1.10 Innocence, Corruption, and the Place of Morality in Human Life

Kant's message is that morality is not about human innocence or natural goodness. Morality plays a crucial role in our lives because we are fundamentally not such innocent or naturally good beings at all, but beings whose deeper nature lies in the capacity to constrain ourselves through reason and according to principle to do those things we often do not have any natural or innocent inclination to do, but can do only if we make ourselves do them. This is no doubt an uncomfortable thing for us to hear, both about morality and about ourselves, because it encroaches on our complacency, by telling us that our life as moral beings is doomed to be an essentially unnatural, uncomfortable, conflict-ridden life.

Virtue for Kant is always a kind of strength—strength of character—which is measured by the resistance it must overcome (MS 6:379–80, 384, 405–9)—a resistance posed by our own moral failings and imperfection: ultimately by the radical propensity to evil we have brought upon ourselves as human

(rational and unsociably sociable) creatures. Moral freedom for us must always be self-government; not a freedom absolutely opposed to coercion, but a freedom even consisting partly in rational self-coercion. Virtue could never be something we might achieve or be given, either by nature or by divine grace, as a kind of supra-moral antinomian state of spontaneous goodness, placing us wholly in the realm of merit or super-merit, in a condition of total moral freedom beyond self-constraint and self-government. He regards the idea of that kind of virtue as not only mistaken, but also arrogant and morally dangerous. To entertain such fantasies, he thinks is to “presume, with proud conceit, like volunteers, not to trouble ourselves about the thought of duty, and, as independent of command, to want to do of our own pleasure what we think we need no command to do” (KpV 5:82). It is, on the contrary, “duty, not merit,” which “must have not only the most determinate influence on the mind, but also . . . the most penetrating influence as well” (KpV 5:157).

This, I think, is why Kant’s readers have strong motives to resist his message, and why moral theories that want to see themselves as naturalistic and healthy-minded are reluctant to take it on board. The common ways of misreading the opening pages of the *Groundwork* is surely in part Kant’s own fault. He is notoriously not an easy or popular writer, and he seems to have been blind to many of the ways this discussion would be likely to be misunderstood. But I don’t regard Kant as entirely responsible for the outlandish ways in which the opening pages of the *Groundwork* have been misinterpreted.

I can’t help thinking that such misunderstandings are sometimes also examples of the sad, sick human tendency to rationalize your dislike of something by exaggerating it in your imagination to the point of absurdity, or even twisting it perversely into what is virtually its diametrical opposite. Thus the recent attempts at a tepid reform of our private health insurance industry, so that it will be a system where fewer people will have to die or face bankruptcy when they get sick, have been depicted by its opponents as a tyrannical imposition of government coercion on every aspect of people’s lives, leading directly to forced euthanasia, gulags, and death camps. With comparable accuracy and, I suspect, analogous self-deceptive motivation, Kant’s thesis that morality is not centrally about natural sympathy but about rational self-constraint is often interpreted as a morbid hatred of all human emotions and a monstrous moral condemnation of all natural generosity.

There is something deeply paradoxical about Kant’s idea that morality is more itself in cases where doing the right thing is hard than in the cases where it is easy, and that those theories which locate the center of morality at the point where goodness is easiest and most natural—theories emphasizing natural virtue or sympathy, for instance—are fundamentally wrong. I think Kant also realized that

this was a paradox, but trusted that his readers would assent to it if it could only be laid before them in starkly clear terms. Two centuries have shown that he either overestimated how compelling his message would be to its intended audience, or underestimated the capacity of his readers to miss it.

Nevertheless, I do find something deeply compelling in the Kantian paradox. I think it tells us something true and important about our modern culture that our deepest standards of self-valuation sometimes put us painfully at odds with ourselves—especially in those cases where morality most shows itself for what it truly is. I believe it would take a major social revolution—whose contours we can barely envision—to make it otherwise. When people object to Kant in the name of a morality that seems “healthier” and more “natural,” less riven by “dualisms” and “constraint,” I suspect these people of living too complacently in a modern capitalist culture they ought to experience as intolerable. They ought to acknowledge to themselves that our culture is deeply incompatible with the possibility of that healthier, more harmonious life they wish to live. Above all, they should stop projecting blame onto Kant (who is only an unwelcome messenger) for the fact that he is right about our moral psychology. They should also stop distracting themselves from the real issues by attributing the features of his moral philosophy and moral psychology that trouble them to the (“two worlds”) metaphysics, which seems to them grotesque and vulnerable to criticism only because they grotesquely exaggerate its (in fact marginal) role in his moral philosophy. Instead, they should credit Kant at this point with profound *social* insight—combined with the profound social discontent he derived directly from Rousseau. It is no doubt the cause of bringing home this message to people that makes me continue the uphill fight of trying to get them to grasp what Kant is actually saying in the early pages of the *Groundwork*.

Kant on Practical Reason

Kant's principal thematic discussion of practical reason occurs in one single place: in the Second Section of the *Groundwork for the Metaphysics of Morals*, where Kant proposes to provide a philosophical derivation of the supreme principle of morality by "follow[ing] and distinctly exhibit[ing] the practical faculty of reason from its universal rules of determination up to where the concept of duty arises from it" (G 4:412). Since Kant here proposes to investigate the "universal rules of determination" of practical reason, we might expect a complete account of practical rationality. But Kant's chief interest here is in *moral* reason, and specifically in deriving the supreme principle of morality. This narrower aim prevents him from providing a truly complete and "universal" theory, by causing him to slight his treatment of the non-moral species of practical reason, to give exaggerated emphasis, in certain respects, to what they have in common with moral reason and also to exaggerate, by contrast, their differences from moral reason. So what we get is an account of practical reason that is perhaps not only incomplete but also skewed by the narrower aim that the account is meant to serve.

2.1 Reason and Reasons

One consequence of this aim is that because his theory of morality in the *Groundwork* began with the concept of duty as moral *necessitation* (constraint in opposition to recalcitrant inclinations), Kant is concerned with practical reason primarily as a faculty of *self-constraint*—that is, a faculty through which we guide our conduct, and even *require* ourselves to do certain things (for reasons). So he is not thinking mainly about cases where action for reasons is easy and natural, harmonizing with empirical desires. Instead, he focuses on cases requiring explicit self-constraint or reflective rational self-supervision.

Some philosophers nowadays distinguish *pro tanto* reasons for doing something—which might be outweighed by countervailing reasons for doing the

opposite—from reasons that strictly require doing the thing. Philosophers also sometimes distinguish *reasons* for doing something (what Christine Korsgaard has recently called “substantive reasons” (Korsgaard, 2009) from “requirements of rationality” (or “rational principles,” such as that to will the end requires you to will the means). Using this distinction, they sometimes argue that (substantive) reasons do not rest on principles, and even question whether requirements of rationality themselves necessarily provide us with any substantive reasons at all (even *pro tanto* reasons) for doing anything (Broome, 2005; Kolodny, 2005).

Kant’s concern with *necessitation* by principles of reason might tempt us to think that he cannot admit there are any *pro tanto* reasons, only reasons that decisively *require* us to do something. This might also lead us to think that he cannot allow for any reasons at all except principles (requirement of rationality), and is therefore committed to the view that such requirements do always furnish us with reasons (indeed, with the only practical reasons there are). Another way Kant’s aim influences his account of practical reason is that since he is looking for the supreme *principle* of morality, he is focused not on the various truths (or facts, or states of affairs) that count for us as practical reasons (in the plural), or on what these reasons may have in common that make them reasons for action, but instead on the general *principles* that underlie these various reasons. This might lead our present-day philosophers of rationality to think that he is interested solely in what some of them call “requirements of rationality,” and not at all in the substantive reasons people have for doing things.

A closer examination of Kant’s account, however, will enable us to avoid such hasty and erroneous conclusions. Kant does think that (substantive) reasons depend on principles, and he definitely does not think of these principles in the way some philosophers now think of “requirements of rationality.” Kant’s concern with necessitation by rational principles gave him few occasions to attend to some of the questions that preoccupy present-day philosophical accounts of practical reason, but I do not think he is committed to any errors or questionable positions on them.

2.2 Reasons and Rational Justification

Yet a third consequence of Kant’s aims in the *Groundwork* discussion is that he seems to have been interested primarily in what we may call (following Francis Hutcheson’s famous distinction) “motivating reasons” (but Hutcheson called them “exciting reasons”), rather than “justifying reasons” (Hutcheson, 1971, pp. 217–18). That is, because Kant is interested in the way certain rational principles might constrain our actions, he is concerned with the *rational motivation* of actions, rather

than in what agents might say to one another (or even to themselves) if asked to give *ex post facto* a *rational justification* for what they are doing. Of course, under the right conditions, the two kinds of reasons might well be systematically related. If what rationally constrains (and motivates) me to do A is that A conforms to a moral principle or duty, then I can presumably justify doing A, at least to myself, perhaps to others as well, by citing that principle and its rationally (morally) constraining force.

In general, however, *justification* is always justification *to* someone. Owing to the relationships in which people may stand to one another, the considerations that justify an action to one person may not serve to justify it to another. For example, if I take an action that negatively affects your interests, I might justify it to you by showing that it violated none of your rights, and also that in acting I showed due concern for your welfare. But what motivated me in performing the action might be something that has nothing to do with you at all, and considerations of your rights and interests might not play any role in the way I justify the action to some third person. However, if I choose an action because I am aware there are certain reasons to do it, then what motivates me ought always to be at least a central part of what I would offer *myself* as my justification for the action.

Some philosophers other than Kant have held that it is indispensable to moral rationality to consider the specific “second-person” claims that one person may have on another. This has been argued powerfully, for instance by Stephen Darwall, who has also drawn some critical conclusions regarding Kant’s own arguments in the *Groundwork* (Darwall, 2006, pp. 213–42). Darwall rightly cites J. G. Fichte, as anticipating much of his argument and as providing an insightful account of the role of second-person reasons, both in our thinking about people’s rights and about moral reasons generally (Fichte, NR 3:29–52). Adam Smith, another great figure in the history of moral philosophy (whose importance, like Fichte’s, has been greatly underestimated), also in effect argued for the indispensability to morality of considering the standpoint of others on the reasons why agents do what they do. On Smith’s account, the “propriety” of an action depends on the attitudes others may take towards it (either with specific concerns in regard to it, or viewing it from a rational impartial standpoint), and toward the feelings or affections from which the action proceeded (Smith, 2000: Part I). Kant’s moral philosophy has been much better studied than either that of Smith or Fichte. On the topic of *justification*, however, and of practical reason in the form of *justifying reasons*, I think these other two great philosophers offer important insights not present in Kant, so that a satisfactory account of these topics would have to pay attention to matters theorized by them but neglected by Kant. Kant was not altogether unaware of these matters, however. I will conclude this chapter by reflecting on the way Kant’s

philosophy does, in a way, take account of the intersubjective dimension essential to all rational justification.

2.3 Practical Reason as Self-constraint by Objective Value

Kant's aim in the *Groundwork* is to work from a consideration of practical reason in general, and its principles, and proceed upwards (or inwards) towards the supreme principle of morality, which in his view is reason's supreme use. For our purposes, however, if we are interested in Kant's general account of practical reason, it is best not to concentrate on (or even to think primarily about) moral rationality until we have to. We will make fewer mistakes if we consider first what Kant says about the other kinds and principles of practical rationality, and especially attend to what Kant thinks all the different kinds of practical rationality have in common.

The main thing they have in common, as we have already seen, is that one primary function of practical reason—the function Kant is most interested in—is *self-constraint*. Principles of reason are principles of rational self-government. This means the motivating force of any kind of reason is *internal* to the faculty of reason itself, and the principles on which it operates. It is not drawn from something outside it, such as an empirical impulse or inclination. Rational motivation itself, therefore, is never furnished by empirical impulses or inclinations. These become relevant to practical reason only when we adopt maxims or set ends based on them, and then ask ourselves questions about these ends and maxims relevant to their rational pursuit—such as what actions would serve to achieve our ends, or what rational priority they should have in relation to other ends, or whether (in light of prudential or moral considerations) we should set these ends at all. Empirical desires become relevant to practical reason only when they engage its principles by being taken up in maxims and the pursuit of ends through purposive choices.

Another point, connected to this one, is that rational principles are always universally valid, valid equally for all rational beings. Their ultimate validity is not dependent on anything (such as contingent desires or the choice of ends) that might distinguish one rational being from another. Of course, the *applicability* (and specific consequences of application) of any rational principle is going to vary a lot from one rational being to another, based on their differing inclinations, knowledge, and contingent situation. For example, whether the principle that willing the end requires me to will the necessary means applies to this action of mine depends on whether it just so happens that I have willed some end and also need

to constrain myself to adopt the necessary means. Similarly, the moral rule that I ought to keep my promise obviously applies to this action only if I have made a promise and this action is required to keep it. But the principles themselves, and the rational grounds for following them, will not vary from one person to another. Kant holds that they belong to the faculty of reason itself, rather than arising out of the data on which this faculty operates. This is what makes them *a priori*, in Kant's usage of the term.

Behind Kant's approach there are some very deep assumptions, no doubt controversial, often not shared by his critics, or even implicitly rejected by them without even realizing what is going on. It may be wise early in this discussion to make some of them explicit. Many familiar (non-Kantian) conceptions of reason either treat desires as implicitly containing basic evaluations, or else as the starting point for a reductive naturalistic account of value. Empirical desires (which Kant would call "inclinations"), are thus taken for granted, at least *ceteris paribus*, by practical reason as the foundation of its activity. The function of practical reason, on this view, is to assist in the satisfaction of desires, by managing their economy, so to speak, and by adjusting their pursuit to the resources available for satisfying them. Naturalistic accounts of practical reason treat it as merely a set of mechanisms that have evolved to perform this function.

Kant, however, distinguishes sharply between empirical desires (or "inclinations") and desires grounded on reason (especially moral reason) (KpV 5:8n; MS 6:211–14). This distinction is, of course, well known, and is often attributed by critics either to an arid intellectualism or to a morbid hatred on Kant's part of the natural and sensual. But this common view rests on a serious misunderstanding. Kant's views at this point instead rest on a rich and interesting theory of human nature, which owes much to currents in the Enlightenment going back at least to Hobbes, but including Mandeville, Diderot, Adam Smith, and above all Jean-Jacques Rousseau. This view holds that human beings are fundamentally and irrationally competitive creatures, whose rivalries serve the natural end of developing their faculties (see Wood, 1999, Chapters 6–9). It therefore treats our empirical desires (or inclinations) with distrust, as arising not merely from our non-rational animal nature, but also from our positively irrational social nature.

On this basis, Kant assumes that empirical desire never automatically provides a reason for acting to satisfy it, and it assigns to our faculty of reason the task of critically evaluating all desires, and objects of desire, a function which grounds the guiding and motivating functions of practical reason. The standards for such evaluation are a set of principles proceeding from the faculty of reason itself (hence their *a priori* status), not to be reduced to natural mechanisms operating on a set of innocent natural desires taken as givens. Reasons must therefore be fundamentally

evaluative (by purely rational objective standards) in a way that separates them from inclinations (*empirical* desires), which contain no evaluation, or none that is not under deep suspicion of being irrational and self-deceptive. The faculty of reason, historically developed through nature's device of social competition, is therefore destined to turn against the very conditions in human nature through which it emerged, and it stands over against empirical desires as a judge and as a constraining authority.

This helps to account for Kant's emphasis on the critical-evaluative and also the self-constraining function of reason, which is crucial, given Kant's picture of human nature, not only in moral reason but also in prudential and even in instrumental reason. It is beyond the scope of this chapter to evaluate Kant's Enlightenment theory of human nature, or to decide how far its assumptions about the relation of reason to evaluation, desire, and motivation might be preferable to the ones that tend to be assumed in non-Kantian accounts of practical reason. My hope is that by bringing into view the tip of the iceberg of Kant's theory of human nature, we can at least begin to appreciate the real grounds for his distinctive approach to the whole topic of practical reason.

But the primary thing to keep in mind here, I believe, is that Kant accepts the idea that there are *objective reasons* for doing or not doing some things, that these are distinct from desires (though they might create desires, or give rational support to already existing desires.) Kant does not hold, as some have thought, that the mere appreciation of reasons or practical principles can or should bring about an action or omission. He thinks human actions always presuppose desires and feelings of certain kinds, though what *rationally* motivates them is never merely a desire. I think Kant's views on this important point have much in common with many of the positions defended at length by Derek Parfit in his wonderful book *On What Matters* (Parfit, 2011)—in fact, more in common than Parfit himself realizes. (But it would take me too far afield to explain this remark further here.)

2.4 Basic Principles and Intermediate Principles

As Kant expounds the practical use of reason, for every basic principle grounding a distinct kind of practical reason, there are intermediate principles, independent of this basic principle, through which the fundamental principle must be applied. For instance, the principle that if you will the end you must will the necessary means has to be applied through "rules of skill" telling you specifically what are the indispensable means to a given end (G 4:415). Likewise, the principle of morality has to be applied through moral rules or duties, which Kant makes the subject of Part I: The Doctrine of Elements in the Doctrine of Virtue (MS 6:417–74). These

intermediate principles are not *deduced* from the rational principle, though it is the rational principle that gives them their practical significance—their motivating (or constraining) force.

For instance, it may be a rule of the medical art that the only way to cure a patient of disease D is to administer medication M (cf. G 4:415). This rule is logically independent of the principle of instrumental reason that tells you that if you will the end you are required to will the means. But that principle, together with the rule of medical art, constrains a physician who wants to cure her patient of disease D to administer medication M. Likewise, in light of the fact that people need the voluntary participation of others in order to be happy, Kant holds that the only way to universalize your maxim of willing your own happiness is to recognize a duty to give them voluntary help. The moral law tells us that we must adopt this maxim of self-love only in a universalized form, so it tells us that we must constrain ourselves (that we have a duty to) give voluntary help to others who need it (MS 6:393, 453). We will see presently that (although Kant himself at times seems confused about some of this), there are analogous intermediate rules (“counsels of prudence”), which are related in a similar way to the basic principle of prudential reason.

Because Kant is considering practical reason chiefly as a faculty of self-constraint (or “necessitation”), he thinks of principles of reason mainly as *imperatives*. “The representation of an objective principle, insofar as it is necessitating for a will, is called a ‘command’ (of reason), and the formula of a command is called an **imperative**” (G 4:413). We probably think of “imperative” first as a grammatical term (characterizing the syntactic form of a certain kind of sentence). But Kant’s conception of an imperative here is part of a philosophical theory of practical reason, so it is not grammatical or linguistic in import. An imperative in his sense is whatever plays the role in deliberation of constraining us by grounding what we do in a rational principle. Kant says that imperatives are expressed by an “ought” (*sollen*) (G 4:413), but this too is obviously not true, if taken as a linguistic remark. Any linguistic formula used in this way counts as an imperative, whatever its grammatical form. Many imperatives (in Kant’s sense) are normally best expressed in indicative sentences from which the auxiliary verb “ought” is wholly absent. As we will see later, for Kant the most natural linguistic form for an *imperative* (in his sense) would be an indicative sentence whose subject is an expression referring to an action and whose predicate is “...is (practically) good” (cf. G 4:413, 414).

Kant famously distinguishes two kinds of imperatives: hypothetical and categorical. The difference between them is that an action represented as necessary by a *hypothetical* imperative presupposes, for its rational necessity, the setting of some end independently of the imperative. A *categorical* imperative rationally necessitates independently of any such pre-given end. It is a conceptual truth for Kant,

however, that all action involves the setting of an end to be produced by the action. In fact, an *action*, properly speaking, is whatever performance we may rationally represent as a means to an end. It is crucial that this last sentence applies every bit as much to actions that follow categorical imperatives as to those that follow hypothetical imperatives. The difference is not whether an action aims at the production of an end (all action for Kant does that), but whether, in relation to the rational requirement to perform the action, the end in question is already given to reason independently of the rational imperative under consideration. In the case of hypothetical imperatives, it is; in the case of categorical imperatives, the imperatives themselves that involve the setting of the end, by specifying the end to be set.

Kant even argues that if there were no ends to be produced that are necessarily connected with the concept of a categorical imperative (“ends that are in their concept duties”), then a categorical imperative itself would be impossible (MS 6:384–5). He holds that there are two general kinds of ends to be produced that are rationally required by categorical imperatives: our own perfection and the happiness of others. To these necessary moral ends, Kant gives the name “duties of virtue” (MS 6:382–4, 385–8). In the case of hypothetical imperatives, however, the imperative always presupposes some end given independently of the imperative—either by contingent, discretionary choice or by prudential reason.

2.5 Instrumental Reason

Kant distinguishes three species of practical reason: (1) technical (or instrumental) reason, which involves a hypothetical imperative with a contingently given end; (2) pragmatic (or prudential) reason, which involves an end necessarily given but dependent on something outside reason itself (namely, the end of happiness); and (3) moral reason, which gives us rational laws in the form of categorical imperatives. Let us consider each in turn—but especially the first two. This is not only because they are de-emphasized in Kant’s own discussion, but also because (perhaps partly for this reason) some of what Kant says about them is problematic and even confused. We will have to interpret, supplement, or perhaps even modify what he actually says about them if we are to arrive at a coherent Kantian account of these kinds of practical reason.

According to Kant, a hypothetical imperative that specifies the means to a contingent or “discretionary” end is called a “problematic” (or “technical”) imperative. (Kant eventually rejected the term “problematic imperative,” holding that in its most natural interpretation, the term even contains a contradiction. In the *Critique of the Power of Judgment*, he insisted that it should be replaced with the term “technical imperative”: KU 5:172–3, see First Introduction to KU, Ak

20:200n). Kant introduces instrumental reason by citing the many “imperatives of skill” that belong to all arts and sciences, and which parents teach their children in the hope that it may enable them to achieve whatever ends they may choose to set later in life (G 4:415). (His examples of such rules are drawn from medicine and from geometrical construction, G 4:415, 418.) Nearly all rules of skill (certainly the medical ones, though not the geometrical ones) are based on merely empirical connections of cause and effect, often connections holding only for the most part or with reasonable probability. Few such rules, if any, specify the *only possible* way to reach their end; thus few can represent an action as strictly *necessary* for its achievement. If an *imperative* is the formula of a command of reason that strictly *necessitates* the actions falling under it—and moreover, a command that is universally valid a priori for all rational beings—then actual rules of skill are virtually never imperatives. So Kant cannot coherently mean to say that “imperatives of skill” are themselves typical examples of technical imperatives. In fact, the content of a rule of skill is never anything but the *theoretical* proposition that a certain result is caused (perhaps even merely for the most part) by a certain kind of action. This point, put together with the idea that all practical reason may be reduced to instrumental reason, is the chief argument sometimes used to provide a pretended defense of the titillating (but silly) Humean paradoxes that there is really no such thing as *practical* reason at all, that I can have no *reason* not to prefer the destruction of the world to the scratching of my finger, and that reason is and ought to be only a slave of the passions (Hume, 1967, pp. 413–16).

Principles of reason are universal in the sense that they apply in the same way to all rational agents. Kant thinks that even rules of skill may be considered “universally valid,” at least in the sense that they propose to tell *anyone who has the end in question* what they must do to achieve it. His intent, however, is apparently not to say that rules of skill are themselves hypothetical imperatives in the sense he means, but rather to suggest that they rest on a single such imperative as their common ground. Thomas Hill calls this general imperative “the Hypothetical Imperative” (Hill, 2002, pp. 172–3). Kant famously formulates it in this way: “Whoever wills the end, also wills (insofar as reason has decisive influence on his actions) the means that are necessary to it that are in his control.” Further, Kant regards this proposition as not only true a priori but as *analytic* (G 4:417). Rules of skill, however (even those, drawn from mathematics, which may be considered a priori) are obviously not analytic:

For in the volition of an object, as my effect, is already thought my causality as an acting cause, i.e. the use of means, and the imperative extracts the concept of actions necessary for this end out of the concept of a volition of this end (to be sure, synthetic propositions belong to determining the means themselves to the proposed aim, but they have nothing

to do with the ground, with making the act of the will actual, but rather with how to make the object actual). That in order to divide a line into equal parts in accordance with a secure principle I must draw two arcs from its endpoints—this mathematics obviously teaches only through synthetic propositions; but that if I know that the specified effect can occur only through such an action, I would also will the action that is required for it—that is an analytic proposition; for to represent something as an effect possible through me in a certain way and to represent myself, in regard to it, acting in that way—those are entirely the same (G 4:417).

If we think that the principle of practical reason involved in technical imperatives (or even hypothetical imperatives generally) is the theoretical proposition that a certain state of affairs can be caused in a certain way, then we have already fundamentally misunderstood the way *practical* reason might bear on the situation. Practical reason is not what is employed in coming to know which action I must perform in order to achieve an end (that is purely a matter for *theoretical* reason). *Practical* reason is rather needed to constrain myself to perform that action *when I already know the theoretical truth* that it is the necessary means to my end that is in my power. (If we think it is not the office of reason, but rather the function of rationally unmotivated desires, sentiments, or passions, to exercise this motivating function, then we are well on the way to the folly of uttering such nonsense as that reason is and ought to be only a slave of the passions.)

Suppose, for instance, that I have set myself an end—for example, to make a good impression on my first day at a new job by showing up on time. Suppose further that I know the theoretical truth that in order to do this, I must get out of bed as soon as my alarm clock goes off. It is not the office of practical reason, in the form of the Hypothetical Imperative, to provide me with this knowledge. That belongs to the theoretical understanding. The job of practical reason is rather to constrain me, in light of this knowledge, to arise on time, overcoming through reason my slothful inclination to stay in bed.

If this point is misunderstood, however, it may lead to a new attempt to reduce practical reason to theoretical reason—this time, to a set of theoretical claims about my psychology. It is often said about instrumental reason that you can avoid rational criticism not only by taking the action necessary for your end, but also by giving up the end. It may be analytic, so the argument goes, that if I will the end, I will the action that counts as an indispensable means, but this analytic proposition holds equally, whether I take the action or give up the end. Therefore, it is claimed, the real question here is still only a theoretical one: namely, whether I do as a matter of fact truly have the end (of making a good impression the first day on the job). If I stay in bed an extra half-hour and show up late, then that shows only (so the argument goes) that I did not really have the end I thought I had; and if I do

get up, then I find out that I did after all have the end—which theoretical truth is demonstrated by the fact that I took the necessary means to it. In either case, no distinctive claim of “practical reason” seems to be involved here at all, only some theoretical truths, one way or another, about my psychology.

This argument, however, gets things entirely wrong about instrumental *reason*. To begin with, it misunderstands the nature of the proposition Kant takes to be analytic. This is not a theoretical but a practical proposition. It is not the proposition that he who (really and truly) wills the end also does in fact will the means, but rather the practical proposition that in performing the rational act of setting the end, I have thereby subjected myself to the practical norm that I *ought* to will the necessary means, thus giving myself a practical (instrumental) *reason* for doing so. The error here may arise from the empiricist assumption that motivation is entirely a matter of states (desires, sentiments, passions) to which we are passive. But setting an end is not something that simply *happens* to you (like being seized by a rationally unmotivated desire). It is something you *do*—it is a rational act, an act of freedom (MS 6:381). Its immediate consequences are not causal but normative.

Non-rational animals, in Kant’s view, cannot set ends, or devise means to them; for them, he thinks, this is all mechanically determined by instinct, perhaps modified by empirical conditioning (MA 8:111–12). With rational agents, he believes, things are different. Setting an end is an act of *volition*, and volition is a normative activity, both in the sense that it can create (subjective) norms (maxims) and in the sense that it is subject to (objective) norms (practical principles of instrumental and prudential reason or moral laws). The concept of an end contains the concept of an action *to be performed* as a means, and therefore lays down the subjective norm that you should perform that action—which norm is, in turn, grounded on the Hypothetical Imperative, the objective (analytical) norm that you should perform whatever action in your power is required as a necessary means to the end.

T.M. Scanlon here follows Broome and Kolodny all the way to the misguided conclusion that the so-called “requirements of rationality”—including those of instrumental reason—are not, properly speaking, normative at all (Scanlon, forthcoming; cf. Broome, 2005; Kolodny, 2005). He thinks that when I choose to perform the action that is a necessary means to my end, my *reason* is not the Hypothetical Imperative (or anything like that), but only the reason I have for pursuing the end. And Scanlon thinks the rationality of my action depends on there actually *being* a genuine or good reason for pursuing the end; it is not sufficient that I *think* there is one. All these claims seem to me erroneous. As a rational agent, I have a *reason* to conform to the Hypothetical Imperative, and this reason is (at least sometimes and in some respects) distinct from my reason for setting the end that makes the Hypothetical Imperative applicable to me. My reason for setting

the end of making a good impression on my first day is quite distinct from my reason for getting out of bed in order to do this.

In adopting the contrary view, Scanlon is moved by the thought that I do not ultimately *have a reason* (or at least a *genuine* or *good* reason) to perform an act necessary for my end if I am mistaken in my belief that I have a good reason to set the end. He may be right about this if by a “good” or “genuine” reason we mean something like a “decisive reason all things considered.” Yet I think we should at least say that as long as I *think* I have a good reason for pursuing the end, this gives me a good or genuine *pro tanto* reason for taking the necessary means to it. Instrumental rationality is made relevant to my present action based on that *pro tanto* reason.

Kant, at any rate, is crystal clear that *instrumental* reason does not require that the end in question should be good in any way. “Whether the end is rational and good is not the question here, but only what one has to do to achieve it” (G 4:415). That is why abandoning my end is not a way of *complying* with the requirements of instrumental reason, but only a way of making them *irrelevant* (to me, at present, regarding the action that was necessary to an end it turns out I do not have). Abandoning an end is a different way of avoiding rational criticism from taking the required means to it. Only the latter is relevant to the question whether I am complying with the demands of *instrumental* reason. As to the former, whether I have a *reason* to abandon my end is obviously not settled by the fact that I have an inclination not to take an action rationally required for it. For instance, my inclination to stay in bed (simply as such and all by itself) constitutes *no reason at all* (not even a *pro tanto* reason) for me to give up the end of making a good impression on my first day on the job. It could become such a reason, but only in a larger context—such as the determination by *prudential* reason that practicing lethargy in the morning is more conducive to my long-term well-being than holding down a job, or the determination by *moral* reason that losing this sleazy job through tardiness would be a fulfillment of my duty. Even if, in light of my overall ends, my desire to stay in bed turned out to be such a reason, that would still have nothing whatever to do with the requirements of *instrumental* reason relative to my end of making a good impression by being on time.

Because Kant’s attention is focused on rational constraint, and the uses of non-moral practical reason that are most analogous to the concept of duty in the case of moral reason, his account of instrumental reason is more limited in scope than seems justified, even by the spirit of what he says about instrumental reason. Kant’s Hypothetical Imperative, for instance, tells us only that we must take any action that is an absolutely indispensable means to an end that we set. Clearly, instrumental reason demands something broader than this—though obviously

far less easy to formulate in the brief space Kant allows himself in the *Groundwork*. To set an end is not merely to lay down a norm to the effect that you should refrain from anything whose omission would absolutely preclude attainment of your end. It also involves some more positive norm to the effect that you ought, *ceteris paribus*, to develop some plan for achieving the end, and either take the feasible steps that belong to the plan, or perhaps modify the plan along the way, so as to work persistently at achieving the end for as long as you continue to have it. But this is only a rough formulation, which it would be hard to make more precise, and only in a very specific context could it yield precise constraints on your actions of the form “You absolutely must do this.” We would surely be selling Kant short if we thought that his conception of instrumental reason must preclude such more vaguely specified and more flexible reasons for action. His very limited discussion of instrumental reason in the *Groundwork* is designed to bring out only some of the basic points, especially those leading to illuminating parallels with moral reason, whose fundamental principle is his sole concern there.

2.6 Prudential Reason

A second kind of practical reason distinguished by Kant is “pragmatic” reason, whose imperatives constrain us in the interest of our own happiness. I think that what Kant says about instrumental reason, if understood along the lines just presented, constitutes a consistent and cogent account. The same cannot obviously be said for his remarks on prudential reason, which require some careful—perhaps even selective—interpretation, if they are to amount to an even minimally coherent story.

Kant is aware of a powerful tendency in the ethical theories of the past to reduce moral reason to prudential reason. He is determined to resist this tendency, and therefore in the *Groundwork* he is motivated to draw as sharp a distinction between prudential and moral reason as he can. For this reason, I think he may exaggerate this distinction at times, even as he himself sees it. This, together with the very limited space he provides himself in discussing prudential reason (again, on his way to his main topic, which is moral reason), creates serious difficulties for anyone trying to develop a defensible Kantian account of practical reason based on this brief discussion. The situation is further complicated by the fact that Kant’s views about human happiness and its pursuit themselves, from the side of empirical psychology or anthropology, contain some insightful and intriguing perplexities and paradoxes, including certain suggestions that the human desire for happiness itself may be in certain respects not entirely rational, or even internally incoherent. These views, precisely because they are of interest in their own right, and challenge

conventional opinions, are bound to complicate further any theory of prudential reason that Kant may propose.

To put the problem briefly, Kant faces a dilemma: on the one side, there is the danger of separating prudential reason so far from moral reason (and so aligning it with instrumental reason) that it ceases to be a distinctive kind of practical reason at all; and on the other side, the danger of recognizing its distinctiveness in a way that opens the door to important questions about prudential reason for which Kant's official account provides no (or no explicit) answers. Resolving this dilemma is an unavoidable task for a reader of the *Groundwork* who is interested in developing a Kantian account of prudential reason for its own sake. This was not a task Kant set himself in the *Groundwork*, and what he says explicitly in that work is clearly not adequate to it.

Kant's treatment of the end of happiness in relation to prudential reason is characterized by three main claims:

The actuality claim: happiness, as distinct from all other empirical ends, may be presupposed a priori without exception as an *actual* end for every finite rational being, an end "belonging to his essence" (G 4:415–16; cf. KpV 5:25; Anth 7:326).

The totality claim: happiness is an idea, framed by the reason and imagination of each individual, of the greatest achievable total satisfaction of all that individual's inclinations (GMA 4:418; cf. G 4:393, 399; KpV 5:124; KU 5:430).

The indeterminacy claim: no individual's idea of happiness is ever determinate or, ultimately, even coherent; it varies over time, often capriciously, and even at any given time no individual can form a precise and consistent idea of what he wishes and wills under the heading of his own happiness (G 4:417–19; cf. KU 5:430–1; Anth 7:233–5).

An important conclusion Kant draws from the indeterminacy claim is that there are, strictly speaking, no imperatives of prudence. There is no action that, for all rational beings, is necessarily required for happiness. At most, there are "counsels (*consilia*)," rules that, if followed, lead, generally speaking and for the most part (but not unexceptionably) to happiness for most people (G 4:418–19). The class of "imperatives of prudence" is necessarily empty.

Kant's attempt to integrate prudence into his conception of practical reason, however, appears to begin with the actuality claim: happiness is not merely a possible end (such as the ends involved in rules of skill), but an end that may be presupposed as actual for all human beings as belonging to their essence (G 4:416; cf. MS 6:387). Therefore, the hypothetical imperative with regard to it is one whose antecedent, so to speak, always holds. This makes imperatives of prudence "assertoric," rather than merely "problematic" (or technical), but not yet "apodictic" (or categorical), because they still rest on a pre-given end (that of happiness) (G 4:415).

From this last point, Kant appears to draw the conclusion that imperatives of prudence—if there were any—would be merely hypothetical imperatives, that prudential reason is simply instrumental reason—but distinguished by the fact that the end in question is in fact always *actual* (not contingent or discretionary).

This official account, however, breaks down in incoherence as soon as we ask even a few simple questions and consider Kant's own explicit answers to them.¹ Human beings have many empirical desires, and many different ends based on them. At any given time, some of these ends, or even a single one of them, is given priority over others and acted upon. Is the actuality claim limited to saying that for every human being, happiness is always in fact one such end among others? Surely there are other ends besides happiness that belong among everyone's ends—the end of having sufficient food and sleep, a comfortable surrounding temperature, and many others too. The priorities among such ends—and their relative priority in relation to other, discretionary ends (ends not all people have) vary greatly from time to time and person to person. If all the actuality claim amounts to is that happiness is one end among others that we all have, then it claims no unique status for happiness among human ends, and therefore the actuality claim does not justify treating “assertoric” (pragmatic or prudential) imperatives any differently from many other hypothetical imperatives. This way of taking the actuality claim, therefore, is incompatible with the idea that prudential reason is a distinctive species of practical reason. It involves either the collapse of prudential reason into instrumental reason, or else it places prudential reason alongside many other possible (or bogus) “species” of practical reason—such as what we might call “alimentary reason,” involving the (also always actual) human end of having enough to eat, or “soporific reason,” involving the (always actual) human end of getting enough sleep, and so forth.

The special and distinctive character of prudential reason therefore seems salvageable only if the actuality claim is understood as involving also what we may call *the unique priority claim*: that among an agent's empirical ends, happiness enjoys a unique priority—perhaps this is what Kant means by saying that it “belongs to [every human being's] essence.” On the basis of Kant's totality claim, it might be held that happiness, as an idea of the greatest sum total of empirical satisfaction, is an end having absolute priority over all other empirical ends, an end to which every other must give way in cases of conflict, or even the single end lending a unique (prudential) reason-giving force to the inclinations that go to make it up.

¹ Patrick Kain provides a stalwart exposition of the official account in Kain, 2003. But he concludes by admitting (Kain, 2003: 250–1) that the account is “incomplete.” I think this rather innocent word underestimates the problems with the official account, which seems to me threatened with hopeless incoherence unless modified in something like the way I am suggesting here.

For instance, from prudential reason we may temporarily postpone even eating or sleeping if we think our overall happiness is best promoted by doing so. Kant says that “as far as our nature as sensible beings is concerned, all that counts is our *happiness*, if this is appraised, as reason especially requires, not in terms of transitory feeling but of the influence this contingency has on our whole existence and our satisfaction with it; [from the standpoint of practical reason as a whole, of course,] happiness is not the only thing that counts,” but this is only because it can be over-ridden by morality (KpV 5:61).

The unique priority claim itself, however, might be taken in either of two senses. First, it might be understood as:

The *empirical priority claim*: that as a matter of fact, human beings do always in fact give their happiness priority over all inclinations or other empirical desires. Or second, it might be taken as:

The *rational (or normative) priority claim*: that practical reason requires them to give it this unique priority, and—at least as far as prudence is concerned (and leaving aside the claims of morality)—practical reason commands that they always prefer their happiness to any other, merely momentary pleasure or partial empirical interest with which it might come into conflict.

Kant’s presentation of the actuality claim as if it were merely a fact about human nature, which could be presupposed to hold of all rational beings, seems to favor the empirical priority version of the actuality claim over the rational priority version. Something like this was even maintained by John Locke, when he said:

All other good however great in reality, or appearance, excites not a man’s *desires*, who looks not on it to make a part of that happiness, wherewith he, in his present thoughts, can satisfy himself. Happiness, under this view, every one constantly pursues, and desires what makes any part of it: Other things, acknowledged to be good, he can look upon without desire; pass by, and be content without (Locke, 1975: II.XXI.§43, p. 259).

The problem with the empirical actuality claim, however, is that it is a claim about empirical human psychology that is plainly *false*. People often overeat, oversleep, or indulge in pleasures that are bad for them in ways they know are counterproductive to their happiness. Locke misses this point because his attention is captivated by the thought that we can think things good without necessarily desiring them (which is true). But it is not true that whatever we desire we desire only as part of our happiness. Moreover, Kant himself knows this very well, and even explains it, on the basis of his indeterminacy claim:

The precept of happiness is for the most part so constituted that it greatly infringes on some inclinations and yet the human being cannot make any determinate and secure concept of the sum of satisfaction of them all, under the name of ‘happiness’; hence it is not to be

wondered at that a single inclination, which is determinate in regard to what it promises and the time in which its satisfaction can be obtained, can outweigh a wavering idea; and the human being, e.g. a person with gout, could choose to enjoy what tastes good and suffer what he must, because in accordance with his reckoning, here at least he has not sacrificed the enjoyment of the present moment through expectations, perhaps groundless, of a happiness that is supposed to lie in health. [So] in this case ... the general inclination to happiness does not determine his will (G 4:399).

So we don't in fact always pursue our happiness consistently, or always prefer it to the satisfaction of particular inclinations. Still less are our inclinations effective in getting us to act, as Locke asserts, only when we understand them to be directed to our happiness. Our inclinations are frequently strong desires that are independent of, even directly opposed to, our end of happiness.

If, therefore, the end of happiness is to have a unique priority sufficient to mark prudential reason off decisively from instrumental reason, then the actuality claim must be understood *normatively*, as the *rational priority claim*. That Kant at least tacitly accepts the priority claim in this sense is also indicated by a passage just quoted, in which he says that with regard to our sensible desires, "all that counts is our *happiness*," adding that this is what "reason especially requires" if happiness is understood not as momentary contentment but the sum total of empirical satisfaction with our "whole existence" (KpV 5:61).

If prudential reason is grounded on the rational priority claim, however, then this has some direct consequences that seem impossible to reconcile with some of the other things Kant says. It also raises some vital questions about prudential reason that he does not answer (and it is not obvious how he would answer). For then it becomes clear that the basic issue of prudential reason is not how to constrain oneself to take the means to an end already given; the issue is rather the rationality of one's ends among themselves, of the priorities among them. Kant's imprudent gout sufferer (G 4:399), for instance, knows perfectly well the *means* to open the liquor cabinet and get at the brandy he desires, and also the *means* to keep his gout from acting up—namely, to leave the liquor cabinet alone and lay off the brandy. What he needs to constrain himself to do, through the exercise of prudential reason, is to give priority to the *end* of his happiness (satisfaction with his life as a whole) over the short-term end of satisfying a momentary impulse the indulgence of which will later leave him much unhappier. So it gets prudential reason basically wrong to focus, as Kant does, on the possible hypothetical imperatives that might be involved in pursuing the end of one's happiness.

Kant is distracted from all this not only by his haste to put prudential reason under the same heading as instrumental reason (the heading of hypothetical, rather than categorical imperatives), but also by the parallel he sees between rules

of skill, on the one hand, and counsels of prudence on the other—precepts of “diet, frugality, politeness, restraint, of which experience teaches that they most promote welfare on the average” (G 4:418). For these look like rules that would lead you to perform actions that are, by and large, means to happiness (whatever you might take your happiness to consist in), in something like the same way that the actions falling under a rule of skill would promote the contingent end of the science or art to which it belongs. Kant’s main contrast between rules of skill and counsels of prudence seems to be based on his indeterminacy claim about happiness, which leads him to the conclusion that there are, strictly speaking, no assertoric imperatives, but only “counsels” that tell us what promotes happiness for most people on average (but allowing for many exceptions). Here he not only seems to be running together the rational imperatives of reason (instrumental and prudential) with the empirical rules falling under them, but also greatly exaggerating the difference between rules of skill and counsels of prudence. It may be that the ends of the arts and sciences are clear and determinate in the way that the end of happiness cannot be for us human beings (though that too is questionable for many arts), but however determinate the ends of an art may be, since there is virtually always more than one way to achieve any end, few rules of skill, if any, apply necessarily and in all cases, so rules of skill are like counsels of prudence in this respect.

We have seen that in the case of instrumental reason, the real imperative of practical reason is not to be identified with rules of skill, but rather with a more basic, universal and *a priori* principle (Hill’s Hypothetical Imperative) that rationally constrains us, once we have set an end, to take the necessary means to it. If we attempt to formulate an analogous principle for prudential reason, it will not concern means–ends relations at all, but rather the relation of priority between the rational end of happiness and all the other contingent ends that are based on inclinations or empirical desires. What is prudent about the gout sufferer’s observing his diet is that it places the end of his happiness ahead of the momentary satisfaction obtainable from an imprudent indulgence; what is prudent about your observing the counsel of politeness is not really that you have chosen the correct means to an already given end, but rather that you discreetly put your long-term self-interest ahead of the momentary impulse to pursue a competing end by blurt-ing out an angry insult at someone who will then become your enemy and go on to do you harm.

In his lectures, Kant distinguishes two tasks of prudential reason: determination of the end (of happiness), and determination of the means to it (VE 27:246). (I am grateful to Grant Rozeboom for calling my attention to this passage, and to the role played in Kant’s conception of prudential reason by the task of “determining the end.”)

The first task shows how the concept of happiness itself is in a way twofold for Kant: It is the concept of the greatest whole of satisfaction of our inclinations, which at the general level is the same for everyone, and it is also a more specific concept of those particular ends that I take to constitute this greatest whole of satisfaction *for me*. The task of determining the end, is the task of moving from the general concept, valid for everyone, to a more specific one, valid for me, that permits of a rational choice of means (the second task). The question naturally arises here whether Kant understands this “totality” to include all my inclinations, or whether I might decide for good reasons to limit my conception of my happiness to the satisfaction of only some of my inclinations—excluding the satisfaction of others from my idea of happiness, even if the inclinations go on existing in me. I think between these two alternatives, Kant clearly opts for the second. For he thinks of happiness as the name for an idea (of reason or imagination) that I make for myself, so that my framing of this idea determines the content of my happiness—it is not determined solely by whatever inclinations I happen to have. In framing my idea of happiness, I must consider not only the strength of my various inclinations, but also my resources for satisfying them. Kant realizes that some ends we may take to be part of our happiness might have unfortunate consequences connected with them that we did not anticipate, so that we might more prudently have excluded these from our idea of our happiness and decided to leave them unsatisfied, or at least to satisfy them only if resources permit, or when bad consequences would not result (see G 4:418–19). Moreover, as is clear from Kant’s indeterminacy claim, he does not think our idea of happiness is ever at any point in time entirely determinate in these respects, and it may also involve various errors, illusions, or incoherences. There might even be rules or maxims of prudential reason to the effect that our idea of happiness should be coherent, that it should include the satisfaction of inclinations whose frustration we know would make us miserable, and that it should not include inclinations for whose satisfaction we have no hope of obtaining the means. Kant never states any such maxims, though it does not seem contrary to the spirit of his conception of prudential reason to do so. The maxims just stated would have to be conditional or *ceteris paribus*, however, since for many of us they might not all be jointly satisfiable.

The rational principle of prudence relates in the first instance to the general concept of happiness, making it a precept of reason to pursue happiness (when-ever this does not conflict with duty). But in consequence it also enjoins the task of making our concept more determinate, and thus an object of practical pursuit (even though, following his indeterminacy claim, Kant thinks we can never do this finally, or perfectly, or even, perhaps, entirely coherently). Only when (and insofar as) the concept of my happiness has been made determinate, can the question arise

of choosing the correct means to it (by way, perhaps, of a hypothetical or technical imperative).

The basic principle of prudential reason, therefore, cannot possibly be a principle that tells us how to take the best means to the end of happiness. Kant holds, as we have seen, that owing to the indeterminacy claim, there are no imperatives of prudence, but only “counsels.” However, there might be a basic *principle of prudence* that could be formulated roughly as follows: “Use your reason, understanding and imagination to form for yourself a more determinate idea of the greatest attainable total satisfaction of your inclinations (under the name of ‘happiness’); and give first priority, among all your non-moral ends, to happiness, preferring it to all other non-moral ends when they conflict with it.” But now we need to ask (as Kant does about both hypothetical and categorical imperatives): “How is this prudential principle possible?” And it is not clear how Kant will answer this question. For we now see that the basic principle of prudential reason is not a hypothetical imperative, telling us how to achieve an already presupposed end; it is a principle rationally directing us to form an end of a certain kind and also to give it unique rational priority over all competing empirical ends. And such a principle is plainly not analytic, as Kant takes the Hypothetical Imperative to be.

At times, Kant suggests that we form the idea of happiness in order to compare our condition as a whole with that of other people, with the aim of judging ourselves superior to them (R 6:27). Arrogance and self-conceit, however, do not look like a rationally creditable basis for prudence. Of course, the principle of morality also says we do have a duty to make the happiness of others our end, which rests on regarding humanity in their person as an end in itself (G 4:430; MS 6:387–8, 393–4). We have no *direct* moral duty to make our own happiness an end, but this is only because Kant thinks we need no direct moral constraint to make happiness an end (MS 6: 388). (It is not clear that Kant is being entirely self-consistent here. For he says that we have a duty to perfect our skills, based on the fact that self-perfection makes us “worthy of our humanity,” even though we “technically practical reason also counsels” us to develop the skills we need (MS 6:387). He might equally have argued that along with prudential reason, moral reason might command us to act prudently—to avoid a bad diet, for instance—as a way of honoring our humanity. This would be a direct and not merely an indirect duty to pursue our happiness, exactly analogous to our duty of self-perfection.)

It might base the *prudential* constraint of reason to give priority to our happiness over other ends of inclination on the fact that we rationally regard our own existence as an end in itself, and prudence constitutes rational regard for our humanity in the same way that the duty to promote the happiness of others constitutes moral regard for their humanity. But such an account of the rational basis of

prudence is nowhere spelled out by Kant, and the general dependence of prudential, or even instrumental reason on moral reason, though it has been suggested by some Kantians, does not seem to be part of his explicit position. If it were, Kantian ethics would be closer to eudaimonism than he seems to want it to be. Christine Korsgaard has presented an interpretation of Kant on which the independence of instrumental reason is called into question, and instrumental reason itself is held to be unintelligible on its own, apart from its foundation in moral reason (Korsgaard, 2007, pp. 215–54). I toyed with a similar idea in Wood, 1999, pp. 55–70, and have also done it again just now regarding prudential reason. But I do not think Kant actually makes any claims of this kind. I don't think his insistence on an *Ungleichheit* (in the sense of unequal status, or even lexical order) between the claims of the three species of practical reason directly commits him to hold that the normativity of instrumental or prudential reason must rest on the normativity of moral reason. On the contrary, it would sooner commit him to the reverse.

In fact, Kant's account of prudential reason seems conspicuously confused, and the confusions, occasioned by his haste to align it with instrumental reason under the heading of a hypothetical imperative, lead him to distort the nature of prudential reason, preventing him even from asking the questions he needs to ask in order to give a defensible account of the rational basis of prudence.

2.7 Moral Reason

Kant says of the three principles of practical reason—instrumental, prudential, and moral—that they are “distinguished by a *difference* (*Ungleichheit*) in the necessitation of the will” (G 4:416). The term *Ungleichheit* here might also be translated “inequality.” (The term in German certainly suggests something comparative, something more than mere non-identity.) That is, we might take Kant as saying that prudential reason trumps or overrides instrumental reason, and moral reason trumps or overrides prudential reason. If we do this, and if we also understand both instrumental and prudential reason as I have already suggested, then we get a fairly complete and coherent Kantian account of practical reason. Reason in its practical use concerns the setting and pursuit of ends. At the most basic level, as instrumental reason it enjoins us to take the means to whatever ends we may have set. But it also imposes rational constraints on which ends we set, commanding us not to set ends based on our empirical desires that are at odds with the end of our happiness, and to limit our pursuit even of our happiness when that is overridden by moral duty.

Moral reason, however, though not based on any pre-given end, does set certain ends, which it is our duty to have. These are our own perfection and the happiness

of others. Kantian theory, however, does not conceive these obligatory ends as admitting of summing or maximizing: it requires that we include various ends falling under the general concepts “our own perfection” and “the happiness of others” among the ends we set, and it forbids us from making our own imperfection or anyone else’s unhappiness (or any part of these) an end (though under many circumstances it might permit us to cause another’s unhappiness as a permissible means or by-product of fulfilling some duty, or even of the permissible pursuit of our own happiness). But it does not require that we place “the greatest happiness” (of any individual) or “the general happiness” (of the whole sentient world) among our ends. (It is consistent, in fact, with regarding these expressions as nonsensical or non-referring, and therefore “the greatest happiness” and “the general happiness” as not possible ends at all.)

2.8 Practical Good

Many contemporary accounts of practical reason are concerned not only with rational principles or “the requirements of rationality,” but even more with the nature of reasons, and with such notions as *pro tanto* reasons. Kant’s account in the *Groundwork* is not concerned with such notions, but I think it would be a mistake to suppose that Kant in any way denies that there are *pro tanto* reasons in the sense used by contemporary philosophers. From what we have seen, in fact, I think we can even begin to develop a Kantian account of what reasons of this kind would consist in.

An important element in developing such an account, I think, is Kant’s interesting formulation of the idea of “practical good”: “The will,” he says, “is a faculty of choosing *only that* which reason, independently of inclination, recognizes as practically necessary, i.e. as good” (G 4:412). “Practical *good*, however, is that which determines the will by representations of reason, hence not from subjective causes but objectively, i.e. from grounds that are valid for every rational being as such” (G 4:413). I take Kant’s notion of practical good to be a somewhat technical one, narrower in application than the way the term “good” is generally used in less precise theoretical contexts (even by Kant himself). Objects or states of affairs we pursue as ends, for instance, are often considered *good*, and Kant himself speaks of them in this way, for instance, in talking about “the highest good” (*summum bonum*) (which includes the satisfaction of our inclinations or our happiness, when we have made ourselves worthy of it) (KrV A810–12/B838,40, KpV 5:110–13; KU 5:448–50). In the strict sense, however, “practical good” refers only to a *way of acting* (KpV 5:60), namely, one that is necessitated by, and grounded, on representations of reason that are objective, valid for all rational beings as such.

The concept of practical good is in place whenever an action is in any way required by an imperative of reason—whether of instrumental, prudential, or moral reason, since all imperatives formulate objectively valid principles. Practical good is what reason requires or grounds, independently of inclination, in the sense that an action's being practically good is a reason for doing it that deserves to prevail over against the mere inclination not to do it. This is true equally of actions that fall under instrumental, prudential and moral reason. It is *not true*, however, that Kant regards actions performed from reason as actions done without interest or desire. This is because he thinks that reason of itself creates an interest in actions and a desire to perform them (G 4:413n; KpV 5:9n; MS 6:211–13). These desires would be, in Thomas Nagel's sense, "motivated" desires (Nagel, 1970, p. 29), and also, in John Rawls' sense "principle-dependent" rather than "object-dependent" desires (Rawls, 2000, pp. 45–8, 151–2).

Practical good may be regarded as the form of *any* practical reason. A fact constitutes a reason for me if, under one of the three basic principles of reason, this fact makes a certain action practically good for me. Thus if I have set an end Z, and it is the case that performing action H is an indispensable means for me to attain Z, then following the principle of instrumental reason (the Hypothetical Imperative), I am constrained by reason to perform H, which makes H practically good (for me under the circumstances), and constitutes a reason for me to perform H. If, as I have suggested, we allow a somewhat more flexible and capacious view of instrumental reason, so that it enjoins not merely the performance of actions indispensably necessary to our ends but also grounds the performance of actions that in one way or another bring our ends about, then the fact that H tends to promote Z would also make H practically good in a looser sense, and give me at least a *pro tanto* reason for performing H. Likewise, if G is a state of mine that belongs to my conception of my happiness, then according to the principle of prudential reason, I have at least a *pro tanto* reason for bringing about G, and also a *pro tanto* reason for performing any action H that is required for G or tends to bring about G. Of course, all reasons furnished by either instrumental or prudential reason are never more than *pro tanto* reasons in the sense that they can be overridden by moral reasons.

Even morality, however, provides me with some *pro tanto* reasons. Morality gives rise, Kant says, to many distinct "grounds of obligation" or "obligating reasons" (*Verpflichtungsgründe, rationes obligandi*), some of which may be stronger than others (MS 6:224). To the extent that any obligating reason can be overridden by a still stronger one, that obligating reason is always a *pro tanto* reason. Morality commands us to make the happiness of others an end, and therefore to include all the elements of every rational being's happiness among our ends. This gives me a

pro tanto reason for contributing to the happiness of any person, or to any part of that happiness. But, of course, the duty to perform any action that contributes to any part of another's happiness is only a wide or meritorious duty (unless special circumstances turn it into more than this). For instance, if you are a stranger to me, and in no special need of my help, then I need not give your happiness any priority among my ends, and there is nothing blamable or contrary to duty in my preferring some portion of my own happiness to it, unless I thereby violate some perfect or strictly owed duty toward you, such as a duty of right.

2.9 “Virtue” Theories of Practical Reason

Kant allows, then, that morality gives rise to different reasons, even *pro tanto* reasons. But he regards them as reasons of a single distinctive kind, based on a rational principle that is fundamental and overriding in relation to all the other reasons we may have to do anything. This view is controversial, and it would be beyond the scope of this chapter to give it anything like a full defense here. But it may help to compare it to fashionable competitors. For example, there are a number of views of practical reason according to which there is no distinctively moral reason that all responsible agents have for doing what they morally ought to do (for doing their “duty,” as Kant would put it). Some of these views, usually associated with the names of Hobbes and Hume, hold that all reasons are either theoretical or merely instrumental. Others, even closer to the real Hume, and associated with what is now called “virtue ethics,” hold that practical reasons, especially those associated with morality, are grounded in the agent's character. (Such a view has been recently articulated with sophistication by Kieran Setiya (2007).)

The main trouble with all these views seems to me that they cannot explain how—or even consistently allow that—a morally bad person, or a person who happens not to have set any moral ends, could ever have a good and compelling reason to do the right thing. If I am such a person, how can I then be expected to do the right thing, or blamed for not doing it? Perhaps this might be done by applying the notion of virtue normatively—though why one should be *blamed* for not acting as the virtuous person acts is still not adequately explained.

These theories say we can take negative attitudes toward morally bad people, or even blame them for moral badness, but it is not clear how they can consider us justified in doing so if by their own lights these people have no reason not to behave in morally bad ways. Suppose I am a callous or cruel person and treat others' rights and welfare with contempt. How can you be justified in blaming me for behaving this way, or in calling me by various bad names (the names of various vices), if you admit I have no reason to behave any other way? If you yourself say

that I have no reason not to behave cruelly or callously, then why should I care what nasty names you call me when I behave in these ways? And where do you get off abusing me with your swear words for doing what you yourself declare that I have no reason not to do? It seems, simply on your own merits, that you (the irrationally abusive moralist) and not I (the callous but wholly reasonable immoralist) are the one open to criticism.

Further, on this theory what counts as virtue has to depend simply on how we feel about people: the “virtuous” person is simply the kind that this person and others like, and there can be no reason (beyond these contingent likings) why anyone should admire such a person, want to be one, or want to act like one. Some people happen to like and admire cruel and arrogant people, who refuse to respect the rights of others and care nothing for their welfare. On this “virtue” theory, it would seem that if more people were like them, then the cruel and arrogant would be the “virtuous,” and there would be no reason to be kind or respectful of the rights of others.

Some philosophers seem to find these consequences acceptable; they exhort us to have the wisdom and modesty to admit that there is, or can be, nothing more to reasons, normativity, and so forth, than what people contingently and empirically happen to like, approve of, and so forth. It seems relevant, on such views that I find their assertions filling me with scorn and indignation, because they think we have no objective reason to care about morality, and seem to regard all moral exhortation as nothing but an attempt to coerce or manipulate one another using irrational feelings. Shouldn’t it count against this kind of “virtue” theory (and its account of reasons, rationality, and so forth) just in its own terms—since it holds that our emotional reactions are the sole sources of value—that many of us regard *it* with disapproval and even contempt?

Present-day virtue theorists often appeal to Aristotle, but if they do so on behalf of the view just mentioned, then they get him wrong in an important way. For Aristotle, if you are a virtuous person, you are virtuous *because* you act for good reasons: you “follow right reason,” as he puts it; and Aristotle repeatedly insists that virtue accords with right reason, *not* the reverse (Aristotle, 1999, pp. 1103b31, 1107a2, 1115b12, 1117a8, 1119a20, 1125b35, 1138a10, b20–34, 1144b2328, 1147b3, b31, 1151a12, a22; cf. *Eudemian Ethics* 1220b19, 1222a8, b5). In other words, it is the fact that you follow right reason that makes you virtuous; it is *not* the case that this or that is a reason (for you) because you happen to be a virtuous person. Reasons come first, virtue comes second. You have to explain what virtue is in terms of reasons; you can’t explain what reasons are in terms of virtue. When you ask what you should do, you are inquiring what there is good reason for you to do *whatever sort of person you might happen to be*. Obviously, in asking what you should do,

and looking for reasons for doing one thing rather than another, one thing you are always asking (directly or indirectly) is *also* what kind of person you should be, which sometimes also includes asking what a virtuous person is like. Reasons are what have to decide that for you, and they can't satisfactorily perform this office if they are treated as mere by-products of a theory of virtue.

Aristotle's "right reason," like Kant's motive of "morality" or "duty" may show itself in a plurality of ways, but it seems to be fundamentally unitary and closely associated (for Aristotle) with the rational part of the soul, which is the supreme human faculty (Aristotle, 1999, pp. 1095a10, 1098a3, 1102b15, 1111b12, 1119b11, 1147b1, 1150b28, 1169a1, a5, 1170b12). It is reason that makes us aware of the practical good, as distinct from the merely pleasant. A "virtue theory" that is authentically Aristotelian would involve a conception of reason that is very close to Kant's.

Let's sum up. From the account we have given, we can see that on Kant's theory of practical reason, the ways actions can be practically good, and therefore the reasons we can have to perform actions, are extremely varied. Inclinations are never simply in and of themselves reasons, but they can give rise to reasons as soon as we set their objects as ends, or include their satisfaction in our idea of our happiness. Facts about how to reach our ends, or the ends of others, are also reasons, sometimes only instrumental reasons, sometimes prudential reasons, sometimes even moral reasons. Moral reason has systematic priority over prudential reason and over hypothetical imperatives about how to reach merely contingent ends, but since some moral duties are imperfect, wide, or meritorious, the moral reasons we have for actions are often merely *pro tanto* reasons, which may often be overridden by prudential or even merely instrumental reasons.

2.10 Rational Justification

It remains to say something about the relation of Kant's account of practical reason to the topic of *justification*. Rational justification, as I said before, is always justification *to* someone, if only to the agent herself whose action is to be justified. A theory of rational justification that might be based on Kant's explicit account of practical reason would probably take the agent's own point of view as primary, and consider justification to others as consisting in getting them to see the action from the agent's point of view, and how it is rationally justified from that point of view.

This "first person" or self-oriented justification seems to work perfectly well when the only kind of practical rationality involved is instrumental or prudential rationality concerning my own ends or my own happiness. But I can also offer another a rational justification for my action by getting the other to understand my end and to understand my action as the necessary, or the best, or at least a rationally

acceptable means to it, or getting the other to see my action, and its end, as contributing to my happiness, according to my idea of it. Here the rational justification, to others as well as myself, consists in showing how the action is practically good according to the Hypothetical Imperative or according to the basic principle of prudential reason. It may even work for some kinds of moral justification, such as the fulfillment of duties to myself, or even the promotion of some other-regarding but impersonal good, such as the utilitarian end of the general happiness (assuming that notion makes sense). Instrumental reason may also have intersubjective uses when people share an end in common, and reason to one another about the best way for each of them to contribute to their shared end.

Prudential reason might work the same way if we could formulate an idea of the collective happiness of a group, and promote it in preference to the partial happiness of its individual members. From a Kantian point of view, utilitarianism looks like a misguided attempt to reduce morality to merely a collective prudential reason. For Kant, however, there is also an essential element of intersubjectivity in rational justification, even when it starts from the standpoint of the agent. For one thing, as we have seen, all the principles of reason (instrumental and prudential as well as moral) are conceived by Kant as objective principles, valid for all rational agents. Even my rational justification of actions—to myself and from my own practical point of view—must rest on principles that are valid even for me only to the extent that they can be seen as valid for others as well. And Kant regards it as an empirical fact that human beings are capable of grasping what is universally valid only by communicating with others and embracing their point of view by understanding their own expression of it.

This is why Kant thinks that “the very existence of reason depends on freedom [of communication]” (KrV:A738/B766) and the use of the judgment of others as a *criterium veritatis externam* (Anth 7:128). Kant therefore formulates the principle of enlightenment—“thinking for oneself” as asking oneself, whenever one is to accept something, whether one could find it feasible to make the ground or rule on which one accepts it into a universal principle of reason (OD 8:146n). And this in turn, is why the maxim of thinking for oneself leads inevitably to the maxim of thinking from the standpoint of everyone else, which one can do only by entering into rational communication with them (KU 5:294–5, cf. Anth 7:200, 228; 25:1480; VL 9:57; Refl 1486 15:715).

Justification on the basis of instrumental and prudential reason, and some forms of moral reason, starts with the agent’s standpoint and becomes justification to others by getting them to adopt that standpoint. Rational justification cannot work in the same way, however, when the issue is another’s rights, interests, or other claims on me. Then things are turned around, and my rational justification of the

action to myself consists in my coming to see my action from the standpoint of the other who is affected by it, and showing how the action respects that other's rights or satisfies the other's legitimate claims or interests. Even in relation to my pursuit of an apparently impersonal good, such as the public interest or the general happiness, justification cannot take place only from my point of view unless we assume that I am entitled (all by myself) to determine the content of these ends, as if I were entitled to choose for others which ends we should pursue in common. Obviously, this is not the right way to think about the ends that are objects of common pursuit by a community of agents. Moral ends in Kant's view, belong to a "realm of ends." An end is an object that a rational being must freely choose for itself: "I can indeed be constrained by others to perform *actions* that are directed as means to an end, but I can never be constrained by others to have an end" (MS 6:381). A "realm" is a systematic unity of ends shared by rational beings, regarded as a universal community standing under a common moral legislation (G 4:433–4; cf. R 6:96–102). In a realm of ends, the laws I am to obey harmonize my ends with a system of ends shared by all rational beings. In that sense, moral justification, following this formulation of the moral law, always has implicit reference to the standpoint of others as well as to the agent's standpoint.

2.11 Mutual Recognition and Community

The primacy of the standpoint of another in Kantian ethics is even clearer when we are talking about moral justification in terms of a person's *rights*, or what we owe to them by duties of respect. "The respect that I have for others or that another can require of me (*observantia aliis praestanda*) is therefore the recognition (*Anerkennung*) of the dignity (*dignitas*) in other human beings, that is, of a worth that has no price, no equivalent for which the object evaluated (*aestimii*) could be exchanged" (MS 6:462; cf. G 4:434–5).

It is not accidental that Kant uses the same term here (*Anerkennung*) that was (at about the same time, yet quite independently) coined by Fichte to capture the essential and reciprocal relation any rational subject must have to other rational subjects, constituting the relation of right (Fichte, 2000, pp. 39–52). But there is something a bit awkward and opaque in Kant's way of formulating the distinction between dignity and price, which focuses on the kind of value something has for an individual agent (and from its own point of view). We are told that something has a *price* if it admits of an equivalent for which it might be rationally exchanged. The *dignity* (of another person, or the moral law, or something whose value is grounded on theirs) is defined only negatively—as that value which cannot be estimated in this way. To determine the price of something is to be given information

that can be used in rational deliberation about what you should do regarding it—how much it is to be valued in your deliberation relative to other things having a price. To be told that something has dignity is to be told only that it has no place in *those* kinds of calculations. Common interpretations of Kant's notion of dignity hold that what has dignity is something having *infinite* value, so to speak, in such calculations. The dignity of humanity would then imply that any act that fails to treat a human being as having infinite value would be forbidden. This invites the objection that this makes it impossible to relate that value to others, and leads naturally to the thought that Kantian ethics must be a system of irrationally strict moral rules, closed to all deliberation based on comparing the consequences of different alternatives.

Let me suggest a way of thinking about dignity that might open the way to a better interpretation. Price is the kind of value that may be used by practical reason in deciding what to do from the standpoint of some agent for whom the factors relevant to a decision may be assigned prices and weighed, by instrumental or prudential reason, relative to them. Dignity, however, is the kind of value that tells the agent that she must not consider the decision only by calculating values from this standpoint, but needs also to combine her standpoint somehow with that of others, and abandon (or at least conditionalize) the whole procedure of deciding what to do by weighing and calculating values from any single standpoint. For a rational agent to appreciate dignity as a value is for that agent to recognize the necessity of shifting from the orientation of providing rational justification to herself (and giving it to others by getting them to see her point of view) to the orientation of including the standpoint of another, or integrating different standpoints into the process of making the decision.

This would mean making the decision not merely by calculating relative to a set of values (prices) attached to factors in it and weighing the overall utility of outcomes, but including in the process also a set of co-operative relationships between different agents who are regarded as parties to it—and to all of whom it must be rationally justified if it is to count as rationally justified to any of them. The idea of the dignity of every rational being also contains the thesis that these co-operative relationships must proceed on terms of the fundamental equality of all persons and mutual respect between them.

There are some grounds for seeing such an interpretation of the positive content of the Kantian notion of dignity as already present—at least implicitly—in Kant's own account of it. The dignity of personhood is the value that for Kant belongs to autonomy, the capacity for self-legislation. The Kantian conception of autonomy might be seen, and often has been seen, in terms of this standard conception of practical rationality, as the giving of moral law to *myself*. But this is not actually

how Kant describes autonomy when he first introduces the idea. He presents it instead as “the idea of *the will of every rational being as a will giving universal law*” (G 4:431). In other words, I legislate to myself to precisely the same extent that I legislate for all other rational beings, and conversely, they legislate to me in exactly the same way that I legislate to them. My will may be regarded as self-legislative precisely to the degree that it is a common rational will shared by all rational beings to the extent that they are ideally rational. The autonomous legislator is not myself as an individual agent, but rather each of us ideally legislating for ourselves and all others. But perhaps the best way of thinking of this ideal legislation is to think of the legislator as the community of rational agents as a whole legislating for all its members through their reciprocal rational interaction. If we think of it that way, then we have transcended the idea of rationality as deliberation from a single agent’s point of view, and opened ourselves to considering intersubjective relations between different agents and different standpoints as constitutive of rationality.

Most traditional theories of practical rationality take the standpoint of a single agent. Even those that want to calculate the good of a collective, or make decisions relative to a common good, such as the general happiness, have operated from the standpoint of a single agent (whether individual or collective in constitution) that simply takes this common good as its own end. The terms of communication and co-operation between agents are then seen as merely instrumental to achieving either the good of individuals or some common good which is the end of the single agent (individual or collective) that is supposed to be calculating the value of utilities and means with a view to achieving it. On this view, rational justification is always justification to that single agent, from its point of view, in achieving whatever ends it has adopted. Kant’s official theory of practical reason as presented in the *Groundwork* looks like a theory of this kind, oriented exclusively to the standpoint of any individual rational being. But if we adopt the interpretation I have suggested of the Kantian notion of dignity, then Kant’s conception of moral value, especially the idea of humanity in the person of every rational being as an end in itself and the dignity of every person as a value surpassing any price, pushes this whole (“first person,” “individualistic,” or “monological”) conception of practical reason to its limits, and even invites us to transcend them.

Kant may thus be seen as the Moses of the moral world, who leads us to the borders of traditional (monological) conceptions of practical reason, and even points us toward what might lie beyond them, but never quite himself sets foot in the promised land.

3

The Independence of Right from Ethics

3.1 Morals, Right and Ethics

Kant's *Metaphysics of Morals* divides "morals" (*Sitten*) into two parts: right (*Recht*) and ethics (*Ethik*). But Kant does not make it clear how right and ethics relate to each other, or how their treatment forms a single whole. It is even unclear, in one basic respect, how the sphere of right can be consistently conceived at all in Kantian terms. Right grounds a set of *duties*—juridical duties or duties of right (*Rechtspflichten*). And all genuine duties for Kant are grounded on categorical imperatives (MS 6:222–3). It is the mark of a categorical imperative, however, that the obligation imposed by it must depend solely on a pure rational incentive, valid for all rational beings as such (G 4:414–15, 427–8). Yet Kant distinguishes the legislation of right from that of ethics by saying that with ethical obligations "the law makes duty the incentive," while the legislation of right "does not include the incentive of duty in the law and so admits of an incentive other than the idea of duty itself" (MS 6:218–19). By this, Kant means that the incentive pertaining to the legislation of right typically involves not a pure rational incentive, but rather incentives provided by external coercion through a public authority. It would apparently follow that duties of right cannot be (or rest on) categorical imperatives at all, and therefore cannot be duties at all, properly speaking. The expression "duty of right" would then contain a *contradictio in adjecto*.

Related to this puzzle, there is also a question concerning the ground of right in general, including the ground of the universal principle of right: "Any action is right if it can coexist with everyone's freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with a universal law" (MS 6:230). One natural way to interpret Kant here is to suppose that the principle of right is somehow derived from the

supreme principle of morality, as that principle is formulated in the *Groundwork* (See Guyer, in Timmons (ed.), 2002, pp. 23–64; Guyer, 2005, Ch. 9; Bernd Ludwig, in Timmons (ed.), 2002, pp. 159–84; for a contrasting view, closer to my own, see Markus Willaschek, in Timmons (ed.), 2002, pp. 65–88 and Katrin Flikschuh, in Denis (ed.), 2010, pp. 51–70). But Kant never tells us how such a derivation might go. And if it is already questionable how the legislation of right can be based on a categorical imperative at all, it has to be equally questionable whether duties of right, or the universal principle of right, can be based on the principle of morality, which is supposed to be a categorical imperative (and was, in the *Groundwork*, even derived from the concept of such an imperative).

Yet another objection to this natural interpretation is that, on closer inspection, the principle of right does not even directly command us to perform actions that are right, or limit our actions to these, but only tells us which actions count as “right.” If in addition to this principle of right, there is also a “law of right” commanding us to perform only actions that are right (MS 6:231), then Kant says explicitly that this law “does not expect, much less demand, that I *myself should* limit my freedom to those conditions just for the sake of this obligation; instead, it says only that freedom is limited to those conditions in conformity with the idea of it and that it may also be limited through deeds (*tätlich*) by others; and it says this as a postulate that is incapable of further proof” (MS 6:231). Kant also adds later that the principle of right is analytic, whereas the principle of ethics is synthetic (MS 6:396). It is far from clear how an analytic principle could need, or even admit of, derivation from a synthetic one. And if the principle of right is a “postulate incapable of further proof,” then that too seems to preclude its being derived from the principle of morality.

All these claims might well be seen as requiring the denial not only that duties of right rest on the categorical imperative, but also that the principle or law of right could be proven through, or rest on, any more fundamental principle at all. Paul Guyer, however, has argued at length (and convincingly), that considered in context, the claims just quoted do not necessarily preclude a deduction of the principle of right from something more fundamental, perhaps even from the principle of morality itself (Guyer, 2005, pp. 203–22). However, there is clearly no discernible “deduction” of the principle of right from the principle of morality in the text of the Doctrine of Right or anywhere else in Kant’s writings. Even Guyer, who believes the principle of right has such a “deduction” from “the concept of morality,” admits that “strictly construed, the claim that Kant’s universal principle of right is not derived from the Categorical Imperative...is correct” (Guyer, 2005, p. 201). And yet if the sphere of right is independent of the moral principle or categorical imperative of ethical duty, then Kant seems to be equally silent on what that independent ground could be.

3.2 Right is Grounded Solely on External Freedom

In my view, the key to finding a path through these difficulties is to recognize that Kant rejects the common idea that the sphere of right, including the philosophy of law and politics, consists merely in an application of general moral principles to the specific circumstances of law or the political state. Kant's theory of right is grounded on one single, very simple and powerful thought, which identifies right with the conditions for protecting what Kant calls "external freedom"—freedom as the independence from constraint by the choice of another (MS 6:237). The thought is this: *because coercion is the restriction of a rational being's external freedom, coercion can be justified only when it is necessary to protect the same external freedom of other rational beings according to universal law. Therefore, the coercion of one rational being by another, involving the substitution of the choice of the coercer for that of the one coerced in regard to the actions of the latter, can never be justified by any end or by any ethical duty.* Coercion cannot even be justified by those ethical duties or ends that might be considered "ends of freedom" in some deeper or richer sense of the word; that is, coercion cannot be justified in order to promote the moral autonomy, or virtue, or the adherence of rational beings to moral ends. It can be justified only to protect the external freedom of rational beings—their freedom even to behave heteronomously and immorally, to violate the moral law of autonomy—as long as they do not thereby infringe the same external freedom of others.

Right does *give rise* to the setting of certain ends—specifically, to the ends of establishing and perfecting the condition of right. The pursuit of those ends, in fact, is the sole legitimate function of the political state and juridical legislation. But right is not grounded on any end; it is grounded solely on the conditions under which all rational beings may have external freedom from coercion according to universal law. This means that we must understand right as *independent* of the moral law or the claims of rational autonomy. There are several ways that those who deny this attempt to ground their denial in Kantian texts or Kantian doctrines. I will now briefly consider four of them.

First, Kant says that the *concept* of duty is taken by the doctrine of right from ethics, and from the moral imperative through which we become acquainted with our freedom (MS 6:239). This derivation of the concept, however, does not make either the legislation of right or duties of right into ethical duties, or bring them under the moral imperative. Indeed, just the opposite is true (MS 6:218–21).

Second, the external freedom protected by right belongs only to free and rational beings, hence only to beings subject to the moral law. But it is not in their moral capacity (their personality), but only in their *humanity*—their capacity to set ends

and choose actions as means to them—that free beings have the innate right to external freedom (MS 6:237).

Third, there is also, as we will see, an ethical incentive for limiting one's actions to what is right, but the duties to which this incentive applies are not ethical duties, but at most “indirectly ethical” (MS 6:221). The legislation of these duties of right falls outside the legislation of ethics, and the principle of morality neither determines the content of right, nor explains why we are required to enter into a condition of right.

Fourth, right also cannot be justified by the way it enables people to *realize* moral freedom or rational agency (see Pippin (in Guyer, 2006, pp. 416–46); and Moyar (in Thorndike, 2011, pp. 137–55)). The way a condition of right enables people to develop and exercise their moral capacities does, of course, exhibit for us some instrumental goods that are achieved in a condition of right. These are great goods, from the standpoint of ethics. In Kant's conception of the human vocation in history, they may even constitute the final end of humanity (KU 5:429–33, Anth 7:321–33); they might even be justifiably called “the realization of freedom.” This end is all very fine from a Kantian point of view; it is even larger and more important, in many ways, than right itself. But we should not let the dazzle of this splendid end distract us from the real questions: “What is the foundation of right? What constitutes the bindingness of duties of right simply as duties of right? What authorizes anyone to coerce their fulfillment?” That foundation of right cannot lie in the way a condition of right serves any end, even an end that can be called “the realization of freedom.” For no rational being can be rightfully coerced to serve any end this being has not chosen, however grand, noble, or important; it can be rightfully coerced only when coercion is necessary to protect the external freedom of others. The sole ground of right is simply the protection of external freedom under universal law.

Distinguishing the conditions under which people may have external freedom in accordance with universal law from all ethical legislation and all ends of morality will enable us to see how right might be a separate sphere of morals (*Sitten*), entirely independent of the sphere of ethics, one that rests on grounds of practical reason that are independent of the categorical imperative that grounds ethical duties, even though the basis of right does have something in common with the basis of ethics, making both of them spheres of morals.

External freedom is the freedom of a person to make choices independently of constraint by the choices of others. The only way for my choices to be completely independent of everyone else's, however, would be for me to have absolute power over all their choices (depriving everyone else of all external freedom). The external freedom of any given person, therefore, must always be limited, since otherwise

no one else could have any external freedom. That is just the point of the universal principle of right: it declares that external freedom, for any given person, is *rightful* freedom only when it can coexist with the same freedom of all others according to universal law (MS 6:230).

The foundation of the entire sphere of right lies in the fact that human beings have good reason to protect rightful freedom. This reason, moreover, has to be independent of *ethical* value, and of the principle of morality grounding *ethical* duties. The problem is to understand this independent ground of rightful freedom, and also how it remains independent of the value that grounds the categorical imperative of morality that Kant derived in the *Groundwork*. Another way to look at this problem is to see it as providing an interpretation and defense of Kant's claim that rightful external freedom—independence from constraint by the choice of another—is “the sole original and innate right belonging to every human being by virtue of his humanity” (MS 6:237). And if the sphere of right is to be independent of that of ethics, this defense must be wholly independent of the *ethical* principle that humanity in persons is an end in itself.

3.3 Ends and the Rational Structure of Action

The solution to this problem, I believe, lies in Kant's conception of the fundamental rational structure of *action*. All action for Kant is based on setting an *end*—an object or state of affairs to be produced. It is the essential characteristic of rational nature to set ends: “Rational nature discriminates itself from the rest in that it sets itself an end” (G 4:437). An *action*, by its concept, is that which lies within the power of the agent, and is chosen by the agent as a *means* to some end (G 4:417). Kant says: “That which serves the will as the objective ground of its self-determination is the *end* . . . By contrast, what contains merely the ground of the possibility of the action whose effect is the end is called the *means*” (G 4:427).

Practical reason applies to action in a variety of ways. First, it applies *instrumentally*: rules of skill tell us the best means to a given end, while a technical imperative of reason, grounding the rational authority of such rules, requires us, on pain of a failure of rationality, to take the indispensable means in our power toward any end we have set (G 4:414–15). Second, reason applies to action *prudentially*: pragmatic or prudential reason counsels us to form an idea of happiness, to make it an end, and to give it rational priority over all other ends of inclination (G 4:415–16). And third, reason applies to action in the form of the *moral law*: the general basis of all ethical duties is the moral law as presented in the Formula of Humanity as End in Itself (FH). An end in itself is not an end we set or an object to be brought about, but something already existing, having a value for whose sake we are required to

do or refrain from certain actions (G 4:427–30). The categorical imperative that grounding our ethical duties commands us to set certain ends—our own perfection and the happiness of others (MS 6:385–8). The moral law also forbids us to set certain ends (for example, the unhappiness of others, pursued for its own sake), and it further forbids us to use certain means to our ends that are inconsistent with treating ourselves or other rational beings as ends in themselves.

All ethical obligations, all categorical imperatives, rest in this way on rational constraints arising out of ends—ends in themselves, rational beings we are required to treat as ends, or ends to be effected that we are required by morality to set. All actions required or forbidden by ethics, or by categorical imperatives, are required or forbidden on account of some end—ultimately, the objective worth of humanity as an end in itself, which we are required to respect in our actions. But the ethical imperative treats actions as required, prohibited, or meritorious on account of their relation to obligatory ends to be produced: our own perfection and the happiness of others (MS 6:385–8). Envious or malicious actions, for example, are forbidden because they make the unhappiness of some person into an end, while actions that perfect our skills or improve our character are meritorious because they further the end of our own perfection, and beneficent actions are meritorious because their end is the happiness of another.

Right, by contrast with ethics, has to do with a class of prescriptions and prohibitions on actions that arise wholly independently of the ends of these actions. Right “has to do only with the formal condition of choice that is to be limited in external relations in accordance with laws of freedom, without regard for any end (the matter of choice)” (MS 6:375). More specifically, right has to do with actions merely insofar as they relate to the freedom of others to choose their own actions in furtherance of ends these others have freely set. Setting an end is an act of freedom. One cannot be compelled (or coerced) to set an end, but one can be coerced to perform actions that are means to ends set by others which are not also one’s own ends. “Another can indeed *coerce* me to *do* something that is not my end (but only a means to another’s end) but not to *make this my end*” (MS 6:381). External freedom consists in choosing actions that are means to ends you have set, while being forced to act in ways that are means to the ends of others is the absence or violation of external freedom. *Right* consists in those rational constraints on our actions that have to do *not* with the ends of those actions but instead with the formal conditions under which your actions are compatible with the external freedom of rational beings in general, according to universal laws that guarantee a like external freedom for all.

The *foundation* of right therefore consists in a rational ground that all rational beings have for requiring the protection of external freedom of all according to

universal law. What could such a ground be? It is this: as a rational being, I necessarily set ends. This implies a rational requirement that I be free to choose the actions by which I pursue those ends. An end is not merely an object I merely *desire* or *wish for*; it is essentially an object I *pursue* through my *actions*, conceived by me as *means* to that end. Human beings—beings with practical reason—are not beings who have their desires satisfied merely by the mechanism of nature (for instance, by the mechanical operation of instincts) or through the beneficent agency of beings other than themselves—such as gods, or robots. For in order to serve the ends of human beings, even gods would have to be prayed to, and robots would have to be programmed and commanded (if only by a remote control device); these actions of praying or commanding would have to be freely chosen as means to the ends the human beings have freely set. Another way of putting it is this: human beings have the privilege—or, if you prefer, they are subject to the curse (Kant recognizes that it can be viewed in both ways)—that whatever good they achieve, and whatever happiness they enjoy, must be the result of their own work: it must arise from their setting an end and their selecting, or perhaps inventing, the means necessary to achieve it, and then applying the means through actions chosen by the human being as such means (cf. I 8:19–20; MA 8:111–12).

It follows that as a rational being, I necessarily will, as far as possible (consistent with other demands of reason), that the actions I perform should serve ends I have set, rather than serving different ends, ends set by others. This is a necessary demand of rational agency, part of its essential structure. It is also a demand entirely independent of whatever particular ends I may have, and even of all the ends I morally ought to have. Moreover, I recognize that other rational beings necessarily make the same rational demand that I do in choosing their actions to further their own freely chosen ends, rather than having their actions constrained to promote ends chosen by others. To recognize a being—whether oneself or another—as a rational agent is to view it as lying under the indispensable rational necessity that it wills to be free, as far as possible, consistent with other demands of reason, to pursue its own ends, rather than being forced to have its actions serve the ends set by others without its freely given co-operation or consent. My demand to be externally free is therefore qualified by the demand of reason that others have the same rational claim on external freedom that I do. No one's external freedom can be unlimited if others are to be externally free as well. This is the ground of the qualification: "consistent with other demands of reason," which must be added to the rational demand for external freedom that each rational being must make.

The ends of morality are objective or binding upon us because they are ends we recognize as rationally valid irrespective of the particular wishes, desires, ends or interests of individuals. The happiness of each individual has a moral claim on

other individuals because it has a value that can be recognized by any rational being. Persons are ends in themselves because they have a value that is not conditional on anyone's empirical wishes or desires but makes an impartial claim on us through a command of reason. Something analogous, but quite distinct, is true of the other constitutive necessity of rational nature I have just identified—namely, its requirement that it be free to choose actions in furtherance of ends set by it rather than having its actions forced to serve ends it does not share.¹

Coercion is a restriction on external freedom: to coerce someone is to force them to act in a way that conforms to your will rather than theirs, and makes their actions serve ends you have set that are not ends they have set. (In moral *self*-constraint our own *rational* will governs our actions, overruling mere inclinations that might have prompted us to act otherwise.) Kant's theory of right is based on the idea that the only way *external* coercion could ever be justified is in the name of protecting external freedom—the very freedom that this coercion annuls, infringes, or limits. A person cannot be justifiably coerced for the sake of any end whatever: neither the happiness of another, nor the happiness of all, nor even the happiness of the person who is coerced, could provide an adequate ground for the coercion of a free and rational being. Such a reason could consist only in the fact that the coercion in question is required in order to protect the external freedom of rational agents in general—*rightful* freedom, external freedom according to universal law. Kant takes it to be an analytic judgment that coercion that protects rightful freedom is itself in accord with right (MS 6:231). “Right and authorization to use coercion therefore mean one and the same thing” (MS 6:232). The fact that right is concerned with external freedom only, and does not depend on ends, is the basic reason why duties of right may be coerced.

Ethical duties, however, are based on ends—on the rational requirement that we set and pursue certain ends that are also duties—which Kant calls “duties of virtue” (MS 6:382–5). These ends belong to the categories of our own perfection and the happiness of others (MS 6:385–8). Because a rational being cannot be externally coerced to set an end, the fulfillment of ethical duties cannot be externally coerced: “Duties of virtue cannot be subject to external legislation simply because they have to do with an end which (or the having of which) is also a duty. No external legislation can bring about one's setting an end for himself, . . . although it may prescribe actions that lead to an end without the subject making it his end” (MS 6:239). But it would be wrong (contrary to right) to try to coerce the fulfillment of

¹ Guyer, 2005, pp. 223–30, has shown that there are both theoretical and practical claims, whose truth is presupposed here, that Kant leaves largely unaddressed. I agree with him, but will not explore the issues now, because they are not relevant to the present inquiry.

an ethical duty by coercing the agent to fulfill an end that agent has not set. The fulfillment of ethical duties must always be through inner *self*-constraint, not external constraint or coercion (MS 6:379). The attempt to derive duties of right from the supreme principle of morality must therefore necessarily fail, since the only duties that can be derived from this principle are duties it would be wrong (contrary to right) to coerce anyone to fulfill.

3.4 Right and Universal Law

It is not a requirement of merely prudential reason, still less of merely instrumental reason, that we should will to be free to choose actions that further our own ends. For this requirement applies not only to our self-interested actions, but also to all our actions as such. As far as reason permits, we will them to be chosen by us to serve our ends, rather than being forced to serve the ends of others. Further, when we take the standpoint of reason, abstracting from all partiality to ourselves, we also see from this perspective that every rational being equally requires that its freedom be protected from destruction, limitation, or usurpation by the choice of another.

And just as morality requires us, when we set our ends and choose means to them, to act on maxims that hold as universal laws from the impartial standpoint of reason, so reason also requires us to recognize every person's equal right to be free from the constraint of the will of others in choosing his actions. This rational requirement of equal freedom consists in having the choice of every rational being protected from external interference to the fullest extent that it can be, consistently with a like freedom of all others according to universal law. The coercion that is consistent with right depends on the claim that everyone's external freedom must be limited in order that all may have external freedom. No one's external freedom should be limited merely for the sake of some end, either of that person or of others. But the external freedom of each of us must be restricted in order that others may be externally free. This is the basis, in Kant's theory of rational action, for right, the supreme principle of right, and all the claims and duties of right. In both ethics and in right, there is a rational constraint imposed by the equal recognition of others, or, as we may also put it, by the requirement that our actions conform to the constraints of universal law. This formal constraint of reason, requiring the recognition of others as having the same claims we do, and limiting permissible actions to those conforming to universal law, is common to duties of right and of ethics. That is what makes these two spheres both spheres of "morals" (*Sitten*).

To put the same point another way: neither ethical duties nor duties of right are based on a selfish calculation to the effect that if I treat others in a certain way, they

will treat me in a similar way. Duties of beneficence are not based on the thought that if I help others when they are in need, then they will (be more likely to?) help me when I am in need, or conversely, that if I do not help them, then they (probably?) won't help me either. Likewise, I do not have a duty to refrain from violating the rightful freedom of others because I think that will make it more likely that they will refrain from violating mine. Both cases involve, rather, an impartial claim of reason grounded on the constraint of universal law.

What the two spheres have in common, therefore, is their subjection to "the categorical imperative, which as such only affirms what obligation is: act upon a maxim that can hold as a universal law" (MS 6:225). I speculate that those who interpret right as grounded on the supreme principle of morality that applies to ethics may do so because they are thinking of the moral law as nothing but the requirement of universalizability. But this is only one aspect of the moral principle—the formal aspect (represented by the Formula of Universal Law or Law of Nature)—which needs to be complemented by formulae that provide ethics with its characteristic ends and motivation (the Formula of Humanity as End in Itself), and that represent ethical legislation as uniting the ends of all rational beings into a system (the Formula of Autonomy or Realm of Ends) (G 4:436). The merely formal aspect of ethics is indeed present in right as well, but only as a constraint constituting the idea of obligation in general, which finds different applications in right and in ethics.

The constraints of right grounded on the freedom of others, and the constraint of universal law that applies to ethics, are therefore analogous constraints, grounded on the categorical imperative that expresses merely "the concept of obligation." But these constraints are not the same, because the application of the concept of obligation (of universal law) is different in the two spheres. The two constraints are also not such that the one constraint could be derived from or be dependent upon the other. The ethical claim that others have on our respect and concern rests on their value as ends in themselves, but their claims of right do not. At most, the ethical value possessed by others as ends in themselves would ground the *ethical* claim that their rights have on us as moral beings. They could never ground the entitlement of one person to have another person coerced or externally constrained to respect their rights. That last claim, to the coercive enforcement of external freedom as independence of another's choice, is one that belongs to right alone. It could not be derived from ethics, since the claims of ethics are never entitled to coercive enforcement. Right therefore rests solely on the claim any person has on any other, from the impartial standpoint of reason, to be left free to choose his actions in furtherance of his own ends, rather than being forced to choose them to serve the ends of someone else.

The basis of right is the protection of external freedom, however it is used, hence not on any of the ends individuals might set. But the condition of right, in which the external freedom of all according to universal law is protected, is a condition into which we may all be coerced to enter (MS 6:256, 312). Preserving and perfecting such a condition can therefore also be seen as itself a kind of end. These are the sole legitimate ends of politics and civil law, the only ends that a ruler or politician as such may rightly seek, and all actions of the public authority may be justified only in relation to these ends. But it is still never the case that coercion can be justified merely because it serves some valuable end or other. Human perfection, happiness, and flourishing, for instance, are valuable ends, but no one may be coerced to promote them just because they are valuable. It is not because the condition of right—its establishment or preservation—may be brought under the general concept of a valuable end that we may be coerced into such a condition or coerced to obey its laws or to fulfill duties of right. Rather, coercion can be justified only because it is needed for the preservation of the freedom of all according to universal law, and the condition of right is nothing except the condition in which this freedom is protected.

External coercion is never justified in the name of any end. Ethical duties always rest on ends. This is why no ethical duty, as such, can ever carry with it a right of external coercive enforcement: forcibly to compel someone to discharge a duty of beneficence, for example, is itself a violation of the right of the person compelled. If duties of right had their basis in the categorical imperative, then they too could not be enforced through external constraint. It is only the rational claim that the external freedom of each person has on every other that grounds duties of right. Ethical duties and incentives, therefore, could not belong to the sphere of right as such. They can neither explain why we have duties of right nor determine the content of those duties. The whole idea that right must somehow be derived from ethics, and that the universal principle of right must either be a version of, or derived from, the categorical imperative, involves a projection onto Kant of the basic error that legal and political philosophy is nothing but an application of *ethics* to the specific circumstances of law and politics. It is an error inimical to human freedom.

3.5 Right as Grounded on Humanity

The ethical claim involves the worth of every person as an end in itself. This involves an ethical worth that belongs to the humanity of every person in Kant's technical sense of the term—humanity as the capacity to set ends and choose means to them (G 4:429; KU 6:431; R 6:27; Anth 7:322–4, 327). The worth of humanity as end in itself provides the end or matter of ethical duty and the motive (*Bewegungsgrund*) for obeying the ethical imperative (G 4:426–9). This end in itself is an *existent*

or *self-sufficient* end (a person having a worth for whose sake we act), as distinct from the ends to be produced that might be set as ends based on this end (for example, the ends of our own perfection or the happiness of others). The basis of the innate right to freedom is also “humanity” in the same technical sense of the term: namely, the capacity to set ends according to reason and choose actions as means to them. Each human being has an innate right to freedom “(independence from being constrained by another’s choice), insofar as it can exist with the freedom of every other in accordance with a universal law, [which belongs] to every human being by virtue of his humanity” (MS 6:237).

This basis, however, is distinct from the worth of humanity as end in itself that grounds ethical duties by providing the motive for obedience to a categorical imperative. For one thing, the worth of humanity as end in itself grounds only non-coercible (ethical) duties, whereas duties of right are subject to external coercion. Also, humanity as end in itself grounds duties on an *end* (humanity as end in itself), whereas claims of right must be independent of every end of action. Ethics places inner (or self-) constraint on the ends-to-be-produced that we adopt (requiring us to include our own perfection and the happiness of others among these ends); but right leaves ends entirely to our free choice, requiring us instead to limit our actions so that they are consistent with the external freedom of others (MS 6:382). Humanity is the ground of right in the sense that to recognize a person as having humanity is to appreciate the rational structure of action as a choice of means to an end set by the agent, and thus the rational requirement that the choice of any rational agent be independent of the constraining will of another in choosing actions as means to one’s own ends. That rational requirement, impartially regarded (hence subject to universal law), is the foundation of right.

3.6 Right as External Constraint, and as Duty

There is even a rather strong sense in which a duty of right does not directly ground my doings or refrainings at all. Recall that the legislation of right does not carry with it duty itself as the legislative incentive. And notice again, that Kant states the universal principle of right not as a proposition that tells us what to do, but simply as a proposition that tells us which actions count as “right.” It tells us, as Kant says, not how we should limit our actions, but how they are limited in accordance with external freedom and hence how they may be rightfully limited by others. A *right* action is one that may not, by standards of right, be coercively prevented, while an action that is *wrong* (*unrecht*), according to those same standards, must be subject to coercive prevention. My duties of right do not, under the legislation of right, exercise rational constraint directly on *me* (this they can do only when

regarded under the legislation of ethics); rather, they specify the use of coercive force that others may rightfully use on me, and it is that coercive force exercised on me that constitutes the bindingness on me of a duty of right. It is in that sense only that duties of right fall under a categorical imperative: they do so for a rational being that sees right actions not merely as actions to which I may be restricted by external compulsion (consistent with the rightful freedom of all under universal law), but also actions to which I must *internally* restrict myself based on the ethical incentive to do only those actions that conform to the concept of universally binding legislation. That is the sense in which the “law of right” can be called a “categorical imperative, which as such only affirms what obligation is: act upon a maxim that can hold as a universal law” (MS 6:225).

The concept of a maxim that can hold as universal law is connected, Kant is claiming, with the concept of obligation, but at a level more abstract than either right or ethics, and this concept applies to the obligations of ethics and right in different ways. In ethics, it is applied to the maxims through which individual rational agents ought to govern their own conduct through the motive of duty and with inner (or self-) constraint. As a rational moral agent, I ought to reject maxims as contrary to duty if they cannot harmonize with universal law (G 4:421), and I ought to adopt any maxim that comprehends within the same rational volition itself as a universal law for all rational beings (G 4:440).

As applied to right, however, this concept concerns not individual self-government, but the conditions under which rational beings may be externally constrained by a commanding will with the power to coerce. Every rational being, as a being that sets ends and chooses actions as means to them, wills that as far as possible, consistent with other demands of reason, its actions should serve its own ends, rather than being constrained to serve the ends of another. As a rational being acting under the idea of obligation as conformity to universal law, it can rationally claim for itself only so much such external freedom as is consistent with a like freedom for all other rational beings according to universal law. This determines the limits of external freedom that are consistent with right. These two applications of the concept of obligation as conformity to universal law are distinct and even disjoint, since the application to ethics concerns inner self-government only, not external constraint, and the latter concerns only what actions may and may not be externally constrained, not the rules for rational self-government.

The universal principle of right, therefore, cannot possibly be based on or derived from the supreme principle of morality—even if what Kant calls “the universal law of right” (MS 6:231) commands categorically. This is because a categorical imperative is one that carries the incentive to its obedience within itself, rather than borrowing the incentive from elsewhere (G 4:414). That is what makes

the moral law governing ethics a principle of inner self-government, rather than a principle of external constraint. Since it belongs to the concept of juridical legislation that the incentive must be borrowed from elsewhere, hence allowing for external coercion as an incentive to obedience, no juridical law or duty of right can, *considered as such*, be a categorical imperative. If, in fact, there is *also* a categorical imperative to comply with duties of right (as Kant holds that there is, in the form of the “universal law of right”), then this must be because there is *also* an *ethical* incentive to comply with the legislation of right. “So while there are many *directly ethical* duties, internal legislation makes the rest of them, one and all [i.e. duties of right], indirectly ethical” (MS 6:220–1).

In other words, there is an ethical incentive, independent of the legislation of right itself, to comply with duties of right. Duties of right are *not* ethical; they must have a foundation independent of the supreme principle of morality. If it is for us a categorical imperative to limit our actions to those that are right (according to the universal principle of right), this is a matter of ethics, not of right. The foundation of that principle, and of duties of right, in fact, the foundation of the entire legislation of right, must be different from and independent of the categorical imperative of ethics. The legislation of right also stands in need of an independent foundation because right makes stronger claims, in one important respect, than any ethical legislation could justify. Specifically, duties of right may be coercively enforced, while the coercive enforcement of ethical duties is always wrong—contrary to right, hence also (if duties of right also make an ethical claim on us) contrary to ethics and to the categorical imperative. Right therefore requires a rational foundation entirely independent of ethics. And it has one, based on the line of reasoning just presented.

This is the only way Kant’s theory of right makes sense. It is the only way that duties of right can be seen as duties whose fulfillment may justifiably be coerced. It is the only way to understand Kant’s claim that ethical legislation includes the incentive of duty, while juridical legislation leaves out the incentive from which duties are to be performed (permitting these incentives to be supplied by external coercive force). It is the only way to see how right and ethics can be two independent parts of morals, while also being subject in common to the general concept of obligation—based on the condition that our maxims must conform to universal law.

3.7 Kantian Right is Sooner Socialist than Libertarian

The basic idea of Kant’s theory of right is that external freedom—not being constrained by the choice of others, not being forced to serve their ends—is an unconditional requirement of reason. It is independent of ethics, and independent of all

ethical values, including the happiness of others, or even of one's own happiness. The exercise of coercion by the state, therefore, can never be justified in the name of anyone's welfare or happiness, not even the welfare or happiness of all—as utilitarian political theories, for example, seem committed to attempt. Coercion can never be justified except in the name of external freedom according to universal law, including the freedom of those coerced.

This basic idea of Kant's theory of right, when stated directly and forcefully, has a tendency to sound like a libertarian slogan. No doubt this is why Kant's theory of right was at one time read by people like Mary Gregor (1963), Robert Nozick (1974) and F. A. Hayek (1976), as supporting their libertarian political ideology. In a recent commentary on the *Doctrine of Right* by B. Sharon Byrd and Joachim Hruschka (2010), this reading is still alive and well.

The libertarian reading of Kant's theory of right, however, represents a decidedly minority position among the experts. Earlier twentieth-century Kantian theories of right—those, for instance, of Hermann Cohen, Paul Natorp, Karl Vorländer, Max Adler, and Julius Ebbinghaus—and also the consensus of more recent treatments of Kantian right—by Wolfgang Kersting (1992), Allen Rosen (1993), Paul Guyer (2000), Leslie Mulholland (1990), Alexander Kaufman (1999), Arthur Ripstein (2009) and myself (Wood, 2008)—have all defended a very different set of conclusions, arguing that Kantian right would sooner result in a social democratic state than in the state friendly to wealth and privilege that is celebrated by libertarians.

Libertarians often claim—thinking that the claim is good Kantianism—that no one's freedom may be restricted by the state for the sake of the welfare of others, or even the welfare of all. And they are right, as far as that one claim goes. But in practice, their ideology advances only the interests of concentrated wealth, capital, and corporate power. That is because their way of thinking refuses to acknowledge that it is equally essential to right that the external freedom of the rich and powerful must be restricted by the state in order to protect the external freedom of the less rich and less powerful. If I must either face destitution or else live only by working for you on your terms, then I am not free to choose how I live. If I have nothing to eat, nothing to wear, no place to live, no protection from injury or disease, then I am profoundly vulnerable to the coercion of others; I do not have a free life in even the most basic sense of the term. It is therefore the responsibility of the commonwealth (the state) to protect every citizen from such a fate: not in the name of welfare, but in the name of freedom. That is *truly* good Kantianism.

Kant does say that the rightful *equality* of citizens, which consists in “each having coercive rights against every other,” is compatible with “the greatest inequality in terms of the quantity and degree of their possessions” (TP 8:291). So he does not

object to economic inequality on any grounds of *equality*. But he takes the rightful *freedom* of each to consist in being able to “seek his happiness in the way that seems good to him, provided he does not infringe upon the freedom of others to strive for a like end which can coexist with the freedom of everyone in accordance with a possible universal law (i.e. does not infringe on the right of another)” (TP 8:290). For this reason, Kant also insists that the rightful *independence* of citizens depends on their having enough property that each is “his own master” (*sui iuris*); no one must be forced to live only “by giving others permission to make use of his powers—and hence that, in the strict sense of the word, he *serves* no one other than the commonwealth” [in its protection of the rightful freedom of all] (TP 8:295). The basic point here was made even earlier by Rousseau, who argued that equality is needed because freedom requires it: “As for wealth, no citizen [should] be so very rich that he can buy another, and none so poor that he is compelled to sell himself” (1997, II.11). As Kant sees it, inequality of possessions, therefore, infringes right less often by infringing the rightful *equality* of citizens than by infringing their rightful *freedom* and *independence*. It is to this that libertarian ideology is systematically (willfully) blind.

Kant thinks that duties of right do not require an assignable individual to whom the duty is owed. He holds that rebellion against a tyrant does no wrong to *him*—since by the injustice of his rule he has forfeited the claim not to be overthrown—but it violates the right, by replacing a law-governed civil society (albeit an imperfect one) with rule by mere force (EF 8:382). He likewise thinks that it is wrong to make a lying declaration—a statement made in a context requiring truthfulness as a duty of right—even to someone whose misconduct has forfeited any right to be told the truth (this is the issue separating Kant from Constant in Kant’s late essay on the right to lie) (VRL 8:426).

Conversely, Kant thinks a person’s right can be violated even if there is no assignable individual who has violated it. Thus if the violation of my right by a particular person can be called “particular injustice,” there can also be a “general injustice” involving a violation of my right, where no assignable individual has wronged me. For instance, a series of economic transactions, none of which in particular involves a violation of anyone’s right, can result in someone’s being in a position in which they are deprived of what is rightfully theirs, and in such a case, it is the responsibility of a law-governed civil society to make arrangements to rectify the general injustice.

One can participate in a general injustice, even if one does no injustice according to laws and institutions. Now if one shows beneficence to a wretch, then one has not given anything to him gratuitously, but has only given what one earlier helped to take from him through the general injustice. For if no one took more of the goods of life than another, there would be no rich and no poor (VE 27:416).

The wealthy, Kant thinks, are the beneficiaries of such injustice, and should be made to pay in order to rectify it. When the government fails to require them to do this, what such people consider beneficence does not really deserve that name:

Having the resources to practice such beneficence as depends on the goods of fortune is, for the most part, a result of certain human beings being favored through the injustice of the government, which introduces an inequality of wealth that makes others need their beneficence. Under such circumstances, does a rich man's help to the needy, on which he so readily prides himself as something meritorious, really deserve to be called beneficence at all? (MS 6:454).

There can be no protection of rightful freedom, Kant holds, without a state authority capable of acting in the name of all, and capable of rectifying general injustice as well as particular injustice. A person cannot be free—their choices independent of wrongful constraint by the ends of others—unless they have a mode of life capable of sustaining that independence. Welfare or happiness, for Kant, is an ethical value, with which right has nothing to do. People should set the happiness of others as an end on ethical grounds, but may not be coerced to promote this end. But rightful freedom is the right of all; it belongs no more to the wealthy and powerful than to anyone else, and the possession of wealth and power give no one the privilege of taking it away from others. When the social or economic system has such a result, it commits a general injustice, which it is the responsibility of civil society to rectify.

Article 25 of the *United Nations Universal Declaration of Human Rights* (1950) holds that “Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control.”

These are rights whose existence is often denied by libertarians, and often represented by libertarians and utilitarians alike under the heading of “welfare.” However, these rights seem to me best justified, as they are by a Kantian theory of right, as necessary conditions of personal freedom and independence. Their violation may sometimes take place through particular injustices, but most often occurs through the general injustice of social customs—for instance, the way in which women are treated in virtually all cultures, or the general injustice of the so-called “free” market (which is often manipulated by the powerful in ways even libertarians pretend to object to—though these ways never seem sufficient to dislodge them from their pious deference to it). A society like ours, that denies a free life to millions upon millions of its citizens through general injustice, subjecting

them through economic vulnerability to the arbitrary choice of others, cannot call itself a free society just because it leaves the wealthy and powerful free to deny others the freedoms that are theirs by right. Instead, it comes close to being, in Kant's technical sense of the term, a barbaric society: ruled by force, where most are deprived of rightful freedom (Anth 7:331). The monstrous and still growing economic inequalities of our society, not only celebrated but even ruthlessly expanded without limit by libertarian ideologues and politicians, now threaten to reduce the vast majority to a condition of servitude to the wealthy and corporations.²

When human rights in Article 25 are justified as necessary conditions for a free life, then those whose wealth and power threatens these rights are not to be viewed as people lacking in "compassion," missing some virtue it would be nice if they had, but is at least acceptable for them to lack. Instead, they are unmasked as what they really are: beneficiaries of wrong, defending their privileges at the expense of the dignity and freedom of others. Their inability to see the world from the standpoint of those they dominate and exploit is relevant only insofar as it plays a role in the self-deceptions through which they depict all limitations on their own unjust power as incursions on rightful freedom. Some politicians who have opposed gay rights, for example, are capable of "evolving" on the issue only when they find they have a son or daughter who is gay—a discovery that apparently enables them for the first time to see the world from a new standpoint. In them, compassion arising from intimate personal acquaintance would appear to be the only available substitute for abstract thought, the capacity for which they appear entirely to lack, at least when it comes to other people's basic human rights.

Those who demand what is rightfully theirs are not beggars asking for hand-outs. They are demanding no more than what properly belongs to them, as human beings and citizens of a just civil society, in which all are entitled to rightful freedom simply in virtue of their humanity. Kant saw these points clearly, and expressed them eloquently, in two memorable unpublished fragments of the 1760s:

Many people take pleasure in doing good actions but consequently do not want to stand under obligations toward others. If one only comes to them submissively, they will do everything; they do not want to subject themselves to the rights of people, but to view them simply as objects of magnanimity. It is not all one under what title I get something. What properly belongs to me must not be accorded me merely as something I beg for (Ak 19:145).

² The title of F. A. Hayek's book *The Road to Serfdom* (Hayek, 1944) was intended to refer to centralized state economic planning, especially in Eastern European socialist countries. Whether or not it was an accurate description of them, the title is an even more accurate (though unintended) description of Hayek's own libertarian economic and political philosophy—hence an accurate description of the philosophy advocated in the book itself. On this point, see Satz, 2010, pp. 174–9.

In our present condition, when general injustice is firmly entrenched, the natural rights of the lowly cease. They are therefore only debtors, the superiors owe them nothing. Therefore, these superiors are called “gracious lords.” But he who needs nothing from them but justice can hold them to their debts and does not need to be submissive (Ak 20:140–1).

It is significant, however, that these quotations are from unpublished notes and lectures (except the one from the casuistical question in the Doctrine of Virtue—and remarks in that context are intended as incitements to the reader’s reflection and practical judgment; they are not necessarily expressions of Kant’s own considered views). For when Kant came to consider the actual economic relations around him, he was largely accepting of existing inequalities, and even of their eighteenth-century political consequences, which very few today would find tolerable.

Kant does argue that the wealthy should be taxed to support the poor, and that this is an obligation of strict right, which is properly a subject of coercion, rather than of voluntary contributions (MS 6:326). For the wealthy owe their existence to their submission to the state, and their right of property becomes a conclusive and peremptory right only through their submission to the general will (MS 6:256–7, 312–13). Therefore, it is left to the decision of the general will how to prevent or rectify general injustice. This allows, at a minimum, for the state to tax the rich in order to provide for the basic needs of the poor. The same considerations could easily justify the taxation of the wealthy to provide for other services essential to a free life: not only: “their most necessary natural needs” (MS 6:326), but all the rights recognized in Article 25 of the *United Nations Universal Declaration*: “food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control.” More broadly still, however, it would seem to require that the state limits inequalities in wealth to the degree necessary to insure that every citizen is *sui iuris*, and able to live freely by his own possessions, not dependent on the will of another.

Kant is willing to accept a contract for services (*location operae*)—of wage labor, as it was found around him (MS 6:285), and he even thinks that household servants who ran away might be forcibly returned to service (MS 6:283). He even defends the occupational and property qualifications that then attached to the right to vote and to participate actively in politics, on the grounds that those who are economically dependent would be unable to act on their own, since they would be beholden to their economic patrons. On this ground, he condemns all servants, peasants, wage-laborers (and, of course, all women) to the status of “passive citizens,” whose rights the state must protect, but who can have no right actively to

participate in political affairs (MS 6:313–15). He is open to the argument that the concept of a “passive citizen” might “contradict the concept of a citizen as such,” but thinks it sufficient if every citizen (he means only male citizens, of course) at least has the opportunity to work his way up to active citizenship (MS 6: 314).

We might, however, grant the insight present in this argument, but use it in the other direction, and infer from the fact that every adult human being has the right to active citizenship that this right is being denied in any society—such as our own—where there are great inequalities of wealth, and consequent constraints on the political choices of many, imposed on them through their economic dependency on others. We might even conclude that wage labor itself, in the form that it actually exists in capitalist society, involves a deprivation of freedom.

On this point, Fichte was more farsighted, insisting that the state should undertake radical redistributive measures and impose tight regulation on the economy, in order to ensure every citizen not only employment but also a “firm estate,” which could be the basis of a life economically independent of the private will of others (NR 3210–215; SL 4:293–8, GH 3:397–421). This contrast between Kant and Fichte has recently been noted, and well treated, by David James (James, 2011, Chapter 3; cf. also Wood, 2008, Chapter 11). The most significant point however, is that these arguments for redistribution and economic regulation are based not on considerations of welfare, but solely on grounds of rightful external freedom.

4

The Moral Politician

4.1 Politics in Kant's Time—and Ours

Kant was not much acquainted with “politics” in our sense of the word.¹ We think of politicians as people who owe their power to being elected (or in some cases, getting other people elected), and who are therefore motivated to gain the approval of enough of the electorate to keep themselves (or those they advise and support) in office. Kant advocated a representative system of government, and he was aware of republican electoral systems in other countries, such as the Netherlands. But he himself always lived under a hereditary despotism, in which high public office was reserved for a hereditary nobility—an arrangement Kant regarded as contrary to the idea of the original contract, and therefore inherently unjust (TP 8:297). To be sure, politicians as Kant knew them needed to keep the approval of enough of the population to ensure the stability of their rule, and they needed to deal with other powerful people whose opposition could threaten their aims or even, in extreme cases, their continuation in office. But they did not have to stand for election or

¹ In this essay, when I speak of “we” and “our” system, I naturally mean only the *American* (the U.S.) political system (the rest of North and South America can always be safely ignored whenever the name “America” is used). We Americans are accustomed to take for granted that everyone in the world must know and care about our politics, while we are entitled to (and generally do) remain contemptuously ignorant of everyone else’s. This is what we proudly refer to as “American exceptionalism.” We consider ourselves (in the words of Ronald Reagan) the “shining city on the hill,” the one lone “beacon of freedom” to all peoples of the world. Experience shows that anyone who is less than eager to agree with us about this will sooner or later live to regret it. For we spend more on our military than is spent on all the militaries of all the rest of the world combined, so that our military is mightier than the combined might of all the militaries of all the nations that have ever existed on earth. That’s why we can confidently expect our extravagant self-congratulations never to be contradicted, and our world domination to be accepted with admiration and gratitude by the entire planet—at least until that point in time (perhaps no longer very far off) when our so greatly admired way of life—which we are always exceedingly reluctant to change or reform, unless in a regressive direction—has made this planet uninhabitable for everyone.

court the favor of some constituency, so that they had less need than we think politicians have for a distinctive political “ideology” or public relations machine, since they did not need to appeal, as our politicians usually do, to what we call a “political base.”

This does not mean, however, that what Kant says about politics is necessarily out of date in relation to “politics” as we know it. The art or craft of dealing with individuals and groups so as to further your ends using them as means (one of the meanings of “prudence” in Kant’s use of this word) has not changed all that much since his time. Our politics, moreover, is in some respects all too much like politics in Kant’s day, even if in our patriotic and self-congratulatory moods we like to pretend it is not. A decent person has much cause to bewail the fact that our politicians are never servants of “the people” as much as they are of the wealthy private individuals or artificial persons (corporations) who finance their campaigns, usually in more or less direct exchange for services rendered. This is neither more nor less than systematic graft and bribery, though even critics of the system usually have better taste than to use those words in speaking of it. What it pleases us to call “democracy” is largely a superficial appearance. Beneath the veneer of popular rule, the electoral system operates largely through the manipulation, by media under the control of the same entrenched interests who pay the politicians, of the foolish and slavish fears, hatreds, and prejudices (cultural, racial, regional) of certain segments of the electorate, which systematically vote against their own rights and interests as well as against those of the vast majority. The privileges pertaining to our system no longer take the form of a hereditary aristocracy, as the privilege of political rule did in Kant’s day, but the content, and the systematic injustices inherent in it, remain very much the same. This means that our politicians need to be successful at fundamentally the same sorts of skills and stratagems as those of which Kant was aware. Hence Kant’s self-appointed task of addressing politicians concerning the moral duties of their profession—as he did most prominently in the Appendix to *Toward Perpetual Peace* (EF 8: 370–86)—has greater lasting interest for us than one would have thought—or might have wished.

4.2 Kant and *Realpolitik*

Kant frames his discussion of politics in the Appendix as a possible conflict between “politics”—regarded as a set of principles of political prudence—and “morals,” as a system of laws that bind us unconditionally or categorically (EF 8:370). This is essentially the same conflict that he takes up under the heading of “theory” vs “practice” in his essay of two years earlier, *On the Common Saying: “That May Be Correct in Theory, but Does Not Work in Practice.”* There, “theory” is characterized

as “a set of principles having a certain generality (*Allgemeinheit*),” while “practice” is “the effecting of an end which is thought as the observance of certain principles of procedure represented in their generality” (TP 8:275). If we suppose the end in question is supplied by the aims and the continued exercise of power on the part of a politician, then political prudence counts as a “practice,” while “morals” could count as the “theory” juxtaposed to it. We see this juxtaposition suggested in the second part of the theory and practice essay, where the arena of the possible conflict of theory and practice is political right (TP 8:289–307).

Kant admits that *subjectively* (“in the self-seeking propensity of human beings”) there can be a conflict (he even at one point calls it an “antinomy”) between right and politics; but he insists there cannot be any conflict *objectively* or “in theory” (EF 8:379). Reason issues moral commands unconditionally, whereas counsels of prudence are always conditional on the adoption of discretionary ends. To the proposition *honesty is the best politics*, which Kant says represents a theory often contradicted in practice, he juxtaposes the “equally theoretical proposition *honesty is better than all politics*,” a principle that is “raised infinitely above all objections and is indeed the indispensable condition of all politics” (EF 8:370). Accordingly, Kant imagines two ways in which a politician might deal with the (possibly oppositional) relation between political prudence and morals. The *moral politician* “takes the principles of political prudence in such a way that they can coexist with morals,” whereas the *political moralist* “frames a morals to suit the statesman’s advantage” (EF 8:372). The “morals” of the latter are doomed to be false and sophistical, and it is only the *moral politician* (Kant is insisting) who has found the correct way of reconciling political prudence with morals.

Those who consider themselves political “realists” are accustomed to condescend to views like Kant’s. They point out that statesmen and politicians face far larger and graver questions than are faced by private persons, and they infer that it is inappropriate to apply to politicians the same standards of “morality” as might be appropriate to ordinary people in the affairs of daily life—as they take Kant to be doing. Carl Schmitt, for instance, holds that the political or public realm has its own criteria, and should not be bound by concepts that can be traced back to those of moral good and evil (Schmitt, 1996). Schmitt seems to conclude that politicians should never be bound by any merely “moral” constraints, as distinct from the practical constraints imposed exclusively by the actual power of others with whom they have to deal in pursuing their political ends.

A more complex view was put forward by Max Weber (1958). He argued that there is a fundamental divide between two kinds of ethics: an ethics of *disposition or ethics of ultimate ends* (*Gesinnungsethik*), and an ethics of *responsibility* (*Verantwortungsethik*). An ethics of disposition is based on principles and ultimate

ends, and the individual feels responsible only for pursuing these strictly and “seeing to it that the flame of pure intentions is not quenched”; the ethics of responsibility, by contrast, is bound mainly by instrumental rationality, takes responsibility for the real-world consequences of actions, and makes no attempt to “dodge the fact that in numerous instances, the attainment of ‘good’ ends [requires] that one must be willing to pay the price of using morally dubious means, or at least dangerous ones” (Weber, 1958, p. 121). The world of the politician, Weber thought, is not one in which the ethics of disposition—of which Kantian ethics is usually thought to be an example—can enjoy full or exclusive authority. Politics is no business for saints, or for finicky and squeamish moralists. The politician’s world is one not merely of conflict but of irrationality and absurdity. In politics, single-minded devotion to an inflexible ethics of disposition will inevitably suffer shipwreck on the reef of modern value pluralism. Politicians must steer a course bounded on all sides by unscrupulous actors of all kinds and by individuals and groups with irreconcilably conflicting values. They need to think strategically, to improvise and compromise, if they are to achieve the best results available (or, perhaps more often, merely to avoid the worst ones). For Weber, the political vocation does not require total abandonment of the ethics of disposition, since politicians themselves bring their own moral goals to their vocation. But politics does require combining the politician’s ethics of disposition with the fundamentally opposed standpoint of an ethics of responsibility. The politician must effect a forcible combination of the two, guided by a strong will and above all by an overriding sense of the responsibility with which the political vocation must bear. “What is decisive is the trained relentlessness in viewing the realities of life, and the ability to face such realities and to measure up to them inwardly... One cannot prescribe to anyone whether he should follow an [ethics of disposition] or an ethics of responsibility, or when the one and when the other” (Weber, 1958, pp. 126–7).

4.3 Right and Ethics

When proponents of such viewpoints consider Kantian ethics, however, they usually entertain only a very simplistic conception of it. They seldom realize how far Kant actually agrees with them in separating the standards applicable to politics from those suited to private life. More importantly, they seldom see where he most deeply disagrees with them, and why. For Kant, “morals” (*Moral* or *Sitten*) is divided into two fundamentally different branches: right (*Recht*) and ethics (*Ethik*). Both involve categorical commands of reason—obligations that must prevail unconditionally—overriding all principles of prudence, and all discretionary ends. *Ethics*, however, provides a doctrine of “morality” as it applies universally to

all human individuals in their private lives. It concerns the virtuous disposition of individuals and their choice of ends. Right, on the other hand, has to do only with the conditions under which the external freedom of all human beings (their independence of coercion by the will of another) can coexist according to universal laws. Creating, maintaining, and perfecting a condition of right is, of course, an end for civil society, and the only ultimate end that politicians as such may legitimately have (their other political ends should always serve this one, directly or indirectly). But right does not concern the promotion of the ends of private citizens, nor does it dictate any ends to them. It concerns only the conditions under which persons may be free to pursue their own ends, consistent with a like lawful freedom of others (MS 6:382).

For just this reason, however, it is only in matters of right, never in matters of ethics, that external coercion is a permissible means of requiring compliance with duties or principles. The creation of a “condition of right”—a condition in which the external freedom of a multitude of people is protected under universal laws—is the sole ground of the legitimacy of a political state, hence the sole ground of any authority rulers, government officials, or politicians may possess. That is the meaning of Kant’s claim that the “moral” proposition “Honesty is better than all politics” is “the condition of all politics.” Without right, a stable republican political order would not be possible, and a legitimate political authority would not be possible.

In the context of the Appendix to *Toward Perpetual Peace*, when Kant speaks of “morals,” he means only right, never ethics. He emphasizes this point repeatedly—at least ten times in the course of a fifteen-page discussion (EF 8:370, 374 (twice), 377 (twice), 378, 379, 381, 383, 384, 386)—so it is truly remarkable that critics habitually overlook it, and usually draw their picture of the “Kantian” position exclusively from what he says about private ethics in the *Groundwork* and other texts that are not about right. When Kant has occasion to speak of ethics at all in the Appendix, he carefully emphasizes its distinctness from right: his main point is always that the politician *as politician* is subject to morals only in the sense of right, not in the sense of ethics (EF 8:381, 386). In other words, when Kant talks about the “moral” principles applicable to politics, he is *not* discussing standards of personal morality or virtue, or anything resting on what Schmitt calls “the moral concepts of good and evil.” Kant is also *not* discussing the individual’s commitment to moral principles and ultimate ends, characteristic of Weber’s “ethics of disposition.” He is referring *only* to those principles and duties that pertain to the *just* exercise of the *coercive power* belonging to the authorities in a *political state*. As I have argued in the previous chapter, Kant even *rejects* the idea that these standards are “moral” in the sense that they belong to or are derived from *ethics*. At the same time, Kant’s

main point in the Appendix is that these standards of right are *not* restricted to standards of political prudence or instrumental rationality—the standards of Weber’s “ethics of responsibility.”

Kant does agree with the realists, however, that politicians are answerable to standards of political prudence that govern the pursuit of their political ends. He praises the “moral politician,” who combines political prudence with the right moral standards, but criticizes the “despotic moralist” who “errs in practice” because, driven by moral enthusiasm, he seeks to implement good measures prematurely and imprudently (EF 8:373). (I am grateful to Jacob Weinrib for calling my attention to this point.) So Kant does think the politician is answerable to a *Verantwortungsethik*, at least up to a point. Standards of right have priority over prudence, because conformity to them is a condition of the rightful exercise of public coercive power over others.

From this we should be able to see that the standards of right that Kant thinks apply to politicians are decisively different in certain specific ways both from the duties of a Weberian *Gesinnungsethik* and from those of a *Verantwortungsethik*. They are drawn solely from the conditions under which politicians may legitimately exercise the power and authority they possess as politicians.

4.4 Political Wrongdoing

Nowadays, when we accuse our politicians of misconduct, what we like to accuse them of, if we can, is *illegality*. We want to catch them *breaking the law* (which some of them do quite conspicuously).² When we can’t show they have acted illegally, we can still accuse them of acting immorally or unethically. This charge can sometimes be made to stick in the court of public opinion, and sometimes even leads to resignation, when the politician finds that the media storm cannot be weathered. But this is a highly contingent matter. Some politicians find they can stick it out, others find they can’t, and it is not knowable in advance which cases will be which. But the charge of immorality more often has to do with personal behavior (especially sexual misconduct) than with specifically political misconduct.

² For instance, when an administration secretly defies explicit congressional restrictions on support to an insurgent movement (financing this support by clandestine sales of weapons to a nation regarded as one of our enemies), or engages in massive secret warrantless surveillance of citizens, or practices illegal detention, torture, or assassination, violating either constitutional restrictions or violating treaties having the force of law, or commits common crimes such as burglary, as when an administration connives to burglarize the headquarters of an opposing party, or the confidential medical records of someone it regards as an enemy, or perjury, as when a president lies under oath about his personal misconduct with an intern.

The standard of right that Kant uses in thinking about wrongdoing on the part of politicians differs in important respects from both the charge of illegality and the charge of immorality. Kant is not the least bit interested in whether the politician is a good or a bad person (virtuous or vicious) as judged by ethical standards. These standards apply to the politician only in private life, and are irrelevant to *political* misconduct. Issues of *illegality* would have also been largely inapplicable in Kant's day to heads of state in a despotic monarchy, where it was taken for granted that the monarch and officials acting in his name are above the law.³

The moral constraints within which Kant's *moral politicians* operate are exclusively principles of *right*. While providing the rational basis for the fundamental legitimacy of all positive laws, they lie equally beyond the standards of ethics and of positive legality. Even the absolute monarch, in Kant's view, owes certain duties of right both to his subjects and (in foreign affairs) to other states with which he deals. The monarch can treat others unjustly, even if no violation of the law is involved—for instance, by imprisoning a subject without cause, or confiscating the property of a subject without justification, or favoring one subject over another in violation of the latter's rights. Positive legislation itself can be unjust, in Kant's view, when it makes one person the slave or bondsman of another, or forbids someone from attaining to a rank to which his talent, industry, and good fortune might take him or otherwise deprives anyone of the free status of being his own master (*sui iuris*) (TP 8:292–5; EF 8: 379; cf. MS 6:241, 31415, 324). Legality, in such cases, would be no defense against the charge of wrong or injustice (*unrecht*)—though Kant's attitude toward unjust laws is generally not that they should be defied (unless what they demand of you is directly contrary to your conscience) but that they have to be tolerated until they can be reformed.

By right, a head of state may also not violate the terms of a treaty into which he has entered with another state, or attack another state with the aim of annexing it (EF 8:35–7, 383–4; MS 6:346). A politician or statesman who does these things, even when prudently protecting his own power and advancing his self-interest,

³ In our day, unconditional claims of presidential (or, in recently expanded versions, more general executive branch) immunity from such accusations usually take the form of appeals to “executive privilege,” or (at their most extreme) the blunt, shameless assertion: “When the president does it, that means it is not illegal.” The claim that “by definition” the president cannot act illegally was (in) famously made by Richard Nixon in his interview with David Frost, April 6, 1977. It was repeated in a speech given by Condoleezza Rice at Stanford University on April 30, 2009, in defense of a wide range of illegalities perpetrated by the Bush administration. Invoking executive privilege became increasingly common under the G. W. Bush administration, where it was extended so as to ensure the silence of virtually anyone who might be in a position to report on the administration's illegal activities. The thought that the president “by definition” acts legally may be seen as unashamed nostalgia, on the part of the executive authority, for the kind of despotic absolutism under which Kant lived.

though he is not misbehaving according to any standards of private morality, is doing wrong (*unrecht*) according to principles of right that are binding categorically or unconditionally. These are principles derived from the concept of a lawful civil society or political order; they ground the authority that any ruler, statesman, or politician may legitimately claim to exercise.

4.5 Can Unjust Rulers be Rightfully Coerced?

Famously (or notoriously), Kant denies that a head of state can be rightfully *coerced* to fulfill these duties of right. All right is in principle enforceable by coercion, but Kant recognizes certain rights as ineligible for coercive enforcement owing to the fact that, under certain circumstances, no power can be in a position rightfully to exercise such coercion. In the sphere of private right, this applies, Kant thinks, to rights of *equity* (Kant regards equity or chancery courts as exceeding the rightful bounds of judicial authority, MS 6:234–5); it also applies to the right of *necessity*, where it cannot be just to punish the violation of a right because the threatened punishment would be incapable of preventing the violation (MS 6:235–6). Kant also thinks it applies to the wrongful acts of a head of state, since “if he could be coerced he would not be the head of state and the sequence of subordination would ascend to infinity” (TP 8:291; cf. MS 6:318–23). Kant therefore rejects Hobbes’s position that the head of state can do no wrong to his subjects, but he also rejects the position of Achenwall and others that it can be just to resist or forcibly depose a ruler who treats his subjects unjustly (TP 8:301). Kant argues that in matters of rightful authority, there has to be a supreme commander or head of state who is the ultimate wielder of coercive power. To suppose there could be rightful coercion of this authority, in his view, would either involve a vicious regress or else simply nullify any effective system of right.

Most of us do not agree with Kant here. We believe that under some conditions, deliberate civil disobedience can be a permissible political tactic. Most of us even think there is a right of revolution against a ruler whose rule is sufficiently unjust. But when we enlist our immediate intuitions (or emotions) against Kant on this point, we ought to acknowledge that it is not easy to offer a well-grounded theory specifying when, and explaining why, disobedience to law or rebellion against a legitimate (though unjust and abusive) authority could be justified. I won’t attempt to re-litigate here all the controversies that surround this issue. But it may be useful to note briefly some easily overlooked features of Kant’s position that make these issues less simple than they are often thought to be.

First, those who condemn Kant’s denial of the right of revolution seldom appreciate the political context in which he first put forward his position, in the essay

on theory and practice. In the year 1793, “theory” was bound to be taken by much of Kant’s audience to refer to the sometimes radical ideas Enlightenment philosophers, while “practice” would refer to the received wisdom of social tradition, to which these theories were often opposed. Kant’s thesis that what is true in theory must also hold in practice was bound to be perceived by many as advocacy of Jacobin political principles, and an apology for the destruction they were at that very moment wreaking in France. Kant’s insistence that rational principles of right cannot countenance the forcible overthrow of an established political order was meant to highlight an aspect of his position that would reassure those who were bound to fear that Kant’s position would incite insurrection, social disorder, terror, and dictatorship.

Second, several scholars have recently pointed out that Kant’s views do permit the forcible overthrow of someone pretending to exercise legitimate authority when the condition is one not merely of despotism, but of *barbarism*. In a barbaric condition, force is being exercised without right; what is needed is not the replacement of one rightful (though unjust) ruler by another, but rather the *establishment* of a condition of right where there is none. The fact that a ruler with coercive powers claimed to be a legitimate head of state is not sufficient to make it true; and if in reality force is being exercised outside any condition of law, the overthrow of barbaric power would on Kantian grounds be not merely permissible but even *required by right* (MS 6:312). The textual basis for this distinction is found in Kant’s *Anthropology*, where Kant distinguishes (1) anarchy (law and freedom without force); (2) barbarism (force without freedom or law); (3) despotism (law and force without freedom) condition, in which right is present deficiently; and (4) republic (force with freedom and law), in which alone there is a true civil constitution (Anth 7:330–1). (See Joerden, 1995; Bernstein, 2008, pp. 57–100; Ripstein, 2009, Chapter 11: Ripstein discusses the case of Nazi Germany in some empirical detail, making a case that it was indeed a condition of barbarism in exactly Kant’s sense of the term).

Third, it has been even less often appreciated that Kant’s views on the duty to obey legitimate authority allows that disobedience is permitted (or even required) when obedience would conflict with “inner morality” or conscience (MS 6:371). But conscience on Kant’s view need not be objectively correct, and even when it is not, we are still required to follow its verdict (MS 6:401, 437–44). If someone believed that he was required in conscience to rebel against the existing head of state—even if this belief is objectively erroneous, as Kant himself would apparently insist that it must be—then Kantian ethics apparently absolves him of the duty to obey, just as it would justify someone in refusing to bear false witness against an innocent person if so commanded by the head of state (for Kant’s enthusiastic praise of the latter refusal, see KpV 5:30, 155–6).

Finally, Kant is persuaded that it is not possible for a head of state to be subject to laws or legal penalties by the argument that such a thing could result only in a vicious infinite regress of higher authorities.⁴ But here he shows a certain lack of political imagination, even in the application of his own principles. He apparently did not consider the possibility that a constitution might provide for the legal removal of any holder of political office, even a chief executive, by way of a process of impeachment and conviction, or of recall by voters. This is certainly a way that a constitution might take account of Kant's own insistence that "the human being is an animal in need of a master"—and the direct corollary he draws from it concerning the construction of a civil constitution: that whoever is placed in a position of command will also be an animal in need of a master (I 8:23). Such legal procedures would not be rebellion; they would instead involve the use of legal and rightful coercion to force a ruler to behave justly (or, at any rate, legally). They would therefore constitute a kind of counterexample—apparently one that could rest firmly on Kant's own principles—to his claim that the executive head of state cannot be coerced to fulfill duties of right.

4.6 The Principles of Publicity

Kant's main treatment of the relation of politics to morals occurs in the Appendix to *Perpetual Peace* (1795). Kant makes no attempt to provide an exhaustive account of the duties of right owed by politicians or of the (surely endlessly many and heterogeneous) ways in which they might commit wrong (*unrecht*) or injustice, some of which have been mentioned earlier. But he does offer two principles, one negative and one positive, by which a politician is supposed to be able to know whether a political maxim or policy might conform to principles of right or violate them. Kant calls these principles of publicity.

The first (or negative) principle of publicity: "All actions relating to the rights of others are wrong (*unrecht*) if their maxim is incompatible with publicity" (EF 8:381). Kant explains:

For a maxim I cannot divulge without thereby defeating my purpose, one that absolutely must be kept secret if it is to succeed and that I cannot publicly acknowledge without unavoidably arousing everyone's opposition to my project, can derive this necessary and

⁴ Kant might plausibly claim that there would be such a regress if we imagined appeals beyond the *sovereign* or ultimate legislative authority (the "general will"), which in a system like Kant's (or Rousseau's) is better seen as a systemic feature of the entire political order than as a particular official or body within it. But the head of state, as Kant conceives it, is the *executive* authority, and there is no reason why the powers of applying the law could not be so arranged that no official who exercises these powers is the sole ultimate exerciser of them, immune to judgment under the laws.

universal, hence a priori foreseeable, resistance of everyone to me only from the injustice with which it threatens everyone (EF 8:381).

Kant's criterion is that a political maxim or policy is known to be unjust when it is possible for the politician to foresee, before implementing it, that its being made public would arouse such public opposition that the aims of the policy would be defeated.

What lies behind this principle is that public right involves enforcement of the general will of the multitude constituting the particular civil society in question (or, in the international cases, the common will represented by a treaty or international agreement). The general will of a political community, for Kant as for Rousseau, is not the will of the majority, but that will which represents the unity of the wills of all citizens: it is a common will that can be represented, according to the idea of an original contract, as the unanimous will of the contracting parties. Even private right (the right of individual property) depends, as a peremptory right, on omnilateral consent, which the public authority of the state must be in a position to give on behalf of all its citizens (MS 6:258–9, 263–4). Therefore, the public authority must stand for what all citizens can be taken to will in common. “The general will of the people has united itself into a society which is to maintain itself perpetually” (MS 6:326). “Public right is a system of laws for a people, . . . who, because they affect one another, need a rightful condition under a will uniting them” (MS 6:311). There is, therefore, a fundamental place in public right for the idea of those laws and policies that tend toward the convergence or union of the ends of all citizens in a common end.

When Kant speaks of the self-defeat being foreseeable “*a priori*,” we must not suppose he means absolutely a priori (as with propositions of pure mathematics, metaphysics, or the fundamental principles of right and ethics). He is instead using “a priori” in the sense in which through experience a man might know a priori that if he undermines his house, then it will collapse; this is known a priori only in the sense that he doesn’t have to wait for the experience of its collapsing to know that it will (KrV B2). In some cases, Kant thinks a policy can be seen in advance to be one that goes against the general will, and will arouse the kind of widespread popular opposition that shows this, just as undermining one’s house can be seen, in conjunction with the laws of physics to justify the conclusion a priori that the house will collapse. Policies that, if publicized, would be self-defeating because the widespread opposition to them shows they are contrary to the general will would be contrary to right, no matter what else might be said in favor of them. Policies that, if publicized, might be unpopular or hard to implement for other reasons are not unjust according to the negative principle of publicity. (We will have more to say about this presently.)

The second (or positive) principle of publicity: “All maxims which need publicity (in order not to fail of their end) harmonize with right and politics combined” (EF 8:386). These policies, Kant thinks, exemplify the sought-for harmony between morals and politics. Again, he provides an explanation: “If this end [sc. of the policy in question] is to be attainable only through publicity, that is, by the removal of all distrust toward the maxims of politics, such maxims must also be in accord with the right of the public, since only in this is the union of the ends of all possible” (EF 8:386). Again, the crucial question is how the policy relates to the general will on which the community is founded. If its success depends on its enlisting the support of the population on behalf of the general will, then it follows that the policy coincides with the general will and is therefore just. The positive principle of publicity should be seen as resting on this basis.

Kant does not pretend that his two principles can decide, regarding any policy whatever that might be brought before it, whether that policy accords with right or is contrary to right. Since the negative principle gives only a sufficient condition for the injustice of a policy, while the positive principle gives only a sufficient condition for its justice, the principles of publicity leave room for many policies, both just and unjust, where the question needs to be decided on other grounds. In this respect, the principles of publicity may be seen as doing in the area of political right what Kant’s Formulas of Universal Law and Law of Nature do in the area of ethics (G 4:421–4). In neither case can they ground specific duties—this is done for ethics using the Formula of Humanity (G 4:429–30); in matters of political right, they are constituted by public right, or “the sum of those laws that need to be promulgated generally in order to bring about a rightful condition” (MS 6:311). Nor do the universalizability formulas in ethics even provide a permissibility test for any and all policies or maxims: they are used only on maxims seen as aiming to justify exceptions to already recognized duties. (For further discussion of these points, see Wood, 2008, pp. 69–74.) In one way, however, the positive principle of publicity is stronger than the universalizability tests in the *Groundwork*, since the positive principle of publicity can be used to show that a policy is just, whereas Kant’s universalizability criterion in the *Groundwork* is used only to show that certain maxims are impermissible; passing these tests apparently constitutes only a necessary condition for permissibility, not a sufficient condition.

4.7 Objections and “Counterexamples”

Despite these restrictions on what they can show, Kant’s claims on behalf of his two principles of publicity are very strong. He regards them as “means for cognizing” what policies are and are not in harmony with right; but he does

not consider them mere heuristic aids, or principles that can be regarded as true only in general and for the most part. He even calls them “transcendental formulas of public right,” and claims they are, “like an axiom, indemonstrably certain” (EF 8:381, 386). These claims for the principles of publicity may appear to be wildly exaggerated, and the first reaction to them of a critical mind is apt to be confidence that it can easily come up with counterexamples to one or both principles of publicity.

A population might consent to—even be enthusiastic about—all sorts of policies that are unwise or unjust; and controversial policies that might be required by justice might be unpopular. Many policies, or at least governmental actions, that are wise, just, and necessary, might also need to be concealed if they are to work. Thus it may be natural to object to Kant’s negative principle, first, that many just policies might fail the negative test because they are too controversial, while second, many unjust policies might pass the negative test; for instance, the policy of unjustly persecuting an unpopular minority within a population would often not need to be concealed. Third, it seems that some such unjust policies might even pass the positive test, since they might be quite widely supported. Fourth, we might worry that the negative principle would rule out clandestine surveillance of terrorist organizations, or other police actions needed to enforce laws that would be easily evaded if the government needed to publicize all its operations.

To take these objections in chiasitic order: policies involving clandestine police actions would be unjust according to the negative principle of publicity only if the reason for their self-defeat were widespread public opposition to them. The fact that criminals or terrorists might elude capture if a policy were made public would not be a reason for regarding it as unjust under this principle. Besides, here we need to distinguish between general policies (to which the principle is meant to apply) and their particular application (to which it is not). It would doubtless be self-defeating, for instance, for a government to announce in advance its intent to raid a specific illegal operation (selling stolen goods, smuggling, and so forth) or to inspect a specific factory for unsafe and illegal working conditions or for adulterating its products contrary to law. Such government practices, however, would not be unjust according to Kant’s principle. For it would not be self-defeating due to widespread popular opposition for the government publicly to announce its intent to use unannounced raids or inspections to enforce these laws. The policy of unannounced factory inspections could even be shown just according to the positive principle of publicity, since publicizing it might be necessary to achieving its direct aim of giving factory owners the intended incentives not to break the law. Some cases of state secrecy are thus clearly compatible with Kantian principles, as long as

the general policies governing them can be made public.⁵ (On all these points, I am indebted to Howard Williams for helping to shape my understanding of Kant's principles of publicity.)

We must be clear that the negative principle of publicity specifies only a necessary or *sine qua non* condition of the justice of a policy, or (to put it contrapositively) a sufficient condition for its injustice. Kant is not saying, therefore (which seems to be presupposed by the second and third objections), that any policy which is popular, or merely not an object of effective public opposition, must *eo ipso* be just or in accordance with right. That would make it all too easy for unscrupulous manipulators of popular ignorance and prejudice, or tyrants able to intimidate the population, to bring morals and politics into harmony. As Kant himself points out: "It cannot be concluded, conversely, that maxims compatible with publicity are on that account also just, for one who has decisive superior power has no need to conceal his aims" (EF 8:385). The policy of unjust persecution of an unpopular minority, for instance, could still be objected to as contrary to right on other grounds, even if it were favored by the majority. Here, publicizing it would in any case not be required for its success, which is what the positive principle demands for the policy to be just. Kant's point in regard to the positive principle is that if a policy needs publicity in order to succeed, then this must be because it spontaneously harmonizes with the universal end of all, and its being made public is required to enlist the support of the population.⁶

Finally: just as the positive principle of publicity is not necessarily satisfied by popular policies, so the negative principle is not necessarily violated by controversial ones. There are many policies, both just and unjust, that would be easier to implement if they remained secret. The mere existence of significant public opposition to a policy is not necessarily a sign that it is contrary to the general will, though Kant thinks that opposition so widespread as to guarantee *a priori* the policy's self-defeat if publicized would be. Sometimes a just policy is harder to implement if publicized because this arouses the opposition of a powerful but unjust or misguided portion of the population that opposes the general will. Kant clearly does not intend his principles to rule out those policies. (Thomas Hurka has suggested that if wealthier Canadians fully realized the redistributive consequences of their single-payer health care system, there would be more political opposition to

⁵ Kant's principles do, however, condemn, for example, the Bush administration's policies of warrantless wiretapping and clandestine surveillance that were kept secret precisely to avoid an anticipated public outcry—the very outcry that actually occurred when they were at last leaked to the public. This is not a counterexample to Kant's principle, but a clear confirmation of its correctness.

⁶ Here we need think only of the disgraceful and unjust laws respecting undocumented immigrants that have recently been passed in our American states of Arizona and Alabama.

it than there is. This may be true, but it does not show that the Canadian health care system violates Kant's negative principle of publicity.)

Further, whether a policy can be known *a priori* to be self-defeating if publicized is not the same question as whether it can be known *a priori* to be easier to implement if kept from the public. It is the former question, not the latter, that the politician is to pose in applying the negative principle of publicity. Applying Kant's principle requires the politician not only to judge whether the policy would be self-defeating if made public, but also to decide whether the self-defeat arises from the policy's violation of the general will or from some other cause. This latter judgment is not a trivial matter, and may require insight and judgment. But it cannot be expected that a moral principle aimed at deciding difficult issues is going to be easy to apply in hard cases. Moral philosophy cannot propose principles that will make it easy even for people of bad judgment and character to decide what to do. Here, as elsewhere in Kantian ethics, good judgment (not reducible to any general rules) is required to apply moral principles correctly (MS 6: 411; Anth 7:199–201).

It may also be thought at this point that it is the decision whether a policy agrees with the general will, and not the question of its self-defeat if made public, that is doing all the work in Kant's principles of publicity. In a sense, of course, this has to be true, since for Kant the ultimate criterion of right of every law or policy is whether it agrees with the general will. But I think Kant believes that this question, taken in the abstract, may be difficult for politicians to answer (or even to make sense of), while the consequences of making the policy public, and how the public reaction will affect its success or failure, may be an easier way for them to raise questions about the relation of the policy to the general will. Kant's principles of publicity are meant to provide politicians with the opportunity to think harder about the way their policies may bear on the common will that grounds the legitimacy of their power.

However, as we will see presently from some of Kant's own application of the negative principle, he thinks it applies mainly in cases where a policy exploits a mutual understanding that all sides take for granted in dealing with one another, but is subverted or undermined by the policy. The terms of such common understandings are obvious instances of the general will. In such a case, if the policy were made public, its self-defeat could be anticipated *a priori* as a result of the way others react to this betrayal of what they have taken for granted. In cases following this pattern, there need be nothing dubious or mysterious about the way the principle applies.⁷ In considering the positive principle of publicity, it is also crucial that the sufficient condition it lays down for a just policy is not merely that its publicity is

⁷ I am indebted to Amanda Greene and Paul Formosa for pressing me on these issues.

needed for it to attain its end, but also that this purposive connection operates by way of creating public confidence in the government, or removing public distrust towards it. This principle supports policies directed towards a union of the ends of all (in other words, carrying out the general will). It would be easy enough to imagine flagrantly unjust policies that attain their wrongful ends through publicity, not by promoting trust and unity but by creating mistrust, fear, and division. A tyrant might publicize unjustly repressive policies in order to intimidate the population; an unscrupulous politician, by publicly inciting one part of the population against another, might more effectively employ the maxim *Divide et impera* (“divide and conquer”), which Kant regards as a conspicuously unjust policy. These maxims are obviously not just according to the positive principle of publicity. It is also not difficult to conceive of similar cases where public announcements might be used to advance Kant’s other examples of unjust maxims: *Fac et excusa* (“Act and then make excuses”), and *Si fecisti, nega* (“If you have done it, deny it”)—or even to find such maxims actually exemplified in the repertoire of our more realistic politicians (EF 8:374–5).

These last two maxims would, of course, also involve public deception—not, perhaps, about the policy itself, but at least about the politician’s intent in following it, or the standards of right by which it is to be judged, or about what has and has not in fact occurred, or other matters. Kant’s positive principle of publicity, however, should obviously not be understood as countenancing maxims as just when they attain their end through publicity by lying to the public. Here again, Kant is appealing to the way people understand the terms of mutual trust on which the legitimacy of a rightful political order must be founded.

4.8 Examples

Kant offers us no examples of policies that can be shown right or just by requiring publicity for their success, saying that he “must leave the further elaboration and discussion of this principle to another occasion” (EF 8:386). He does provide discussion of several examples of the application of the negative principle. Not all these discussions are equally satisfactory, but a brief review of them may help us to understand the negative principle better.

The right of a state: the maxim of rebellion: Kant’s example of a maxim pertaining to the right of a state does not concern the behavior of those who hold authoritative political power, but at most of those who aspire to do so. This is the maxim of civil rebellion as a way of ridding a people of an oppressive or tyrannical ruler. Kant concedes that if the ruler is unjust, then no wrong is done to him through deposing him. But he insists that rebellion is an unjust means to achieve

its end, and that this can be shown by its violation of the negative principle of publicity: “The wrongfulness of rebellion is therefore clear from this: that the maxim of rebellion, if one publicly acknowledged it as one’s maxim, would make one’s own purpose impossible” (EF 8:382).

This is far from self-evident, however. No doubt it could be known a priori that publicizing the maxim of rebellion would rouse the tyrant to defend himself; but the same could equally be known a priori for any conceivable *just* means of deposing him. That would also not involve self-defeat due to widespread public opposition to the policy—or it would do so only if the existing (tyrannical) regime had more widespread popular support than the revolutionary cause. But in that case, the rebellion would be merely an unpopular *coup d’état* or putsch, which might well be contrary to right, even against a tyrant. The *Communist Manifesto* proclaims in its concluding paragraph:

The Communists disdain to conceal their views and aims. They openly declare that their ends can be attained only by the forcible overthrow of all existing social conditions. Let the ruling classes tremble at a Communistic revolution. The proletarians have nothing to lose but their chains. They have a world to win. WORKING MEN OF ALL COUNTRIES, UNITE! (Marx and Engels, CW 6:518).

Marx and Engels sought as wide a publication as possible for this greatest of all revolutionary pamphlets. It is true that soon after publishing these words, Marx was forced to leave the continent of Europe for a lifelong and often painful sojourn in England. But could Marx have known this a priori? And did the *Manifesto*’s public announcement directly frustrate the political aims Marx and Engels had in publishing it? To both questions, the obvious answer is: “No.”

Other cases are even worse for Kant’s argument. It is hard to see how Kant’s argument could apply to genuine popular uprisings, such as the recent ones against Ben Ali in Tunisia or Mubarak in Egypt. It might even be argued that these insurrections even required publicity to achieve the overthrow of Ben Ali and Mubarak, in which case it seems evident that these policies of rebellion were necessarily just, according to Kant’s positive principle of publicity.

Ernest Weinrib has proposed to me in discussion that we interpret Kant’s argument here in the following way. In an already constituted civil order, the laws and executive authority may be presumed to have the support of the general will. Therefore, it can be known a priori that the policy of attempting their overthrow will excite the kind of opposition that shows this policy to be contrary to the general will, and this is why Kant thinks it can be known to be unjust according to the negative principle of publicity. Such an argument would seem to rest on the notion that in any existing political order, there is a general understanding that the government is to be obeyed, and that the maxim of rebellion attempts to exploit

this understanding in order to overthrow or disobey. But that invites the observation that sometimes it seems actually to happen—as in Tunisia and Egypt—that an existing political regime loses the support of the general will and can no longer be seen as its embodiment. In that case, the general understanding—which is plain from the scope of public demonstrations—may have exactly the opposite import. Kant may have formidable arguments against the right of revolution; this argument from the negative principle of publicity is not one of them.

The right of nations: Kant's three examples relating to international right were apparently drawn from Christian Garve's treatise *On the Combination of Morals with Politics* (1788) (EF 8:385n). Kant graciously praises Garve as a scholar (as is his habit in speaking of Garve), but he clearly regards Garve's tolerant attitude toward these maxims (about which he is reluctant to judge either way) as far too permissive.

Kant considers three policies concerning the external relations of states that a politician might consider prudent, but (according to Kant) can be shown to be contrary to right using the negative principle of publicity:

a) A state reneges on an international agreement

Kant imagines a head of state who has promised something to another state (some service, or the cession of territory, or a subsidy, etc.). The head of state intends to back out of the agreement, reasoning that if he made the promise as a sovereign power, owing nothing to his own state internally, he can release himself from the promise in his role as an official or servant of the state, whose duty to it overrides his obligation to the other state. Kant argues that if this policy were made public, other states would not trust him and might even unite against him to oppose his ends, thus rendering the policy self-defeating.

b) Small states unite in a preventive war against a more powerful neighboring state

The smaller states, knowing that the larger state has the power to oppress them, fear that it may also acquire the will to do so, even though it has done no injury to any of them. They join in attacking it, to pre-empt the threat even before it has a chance to materialize. Kant argues that if they made their intention public, this would incite the very oppressive behavior they fear, and thus the policy would be self-defeating.

c) A larger state annexes a smaller one in order to make its territory more secure

A smaller state's territory separates the territories of the larger one, which it thinks will be more secure if it possesses all the territories continuously. It forcibly

subjugates the smaller state, and considers itself justified in doing so, on grounds of self-preservation. Kant argues that if it made this maxim public, the smaller state would form a defensive alliance with other neighboring states, not only denying the larger state its territory, but even increasing the danger the larger state aims to prevent.

Kant's arguments concerning these cases of international right, unlike the one about the maxim of civil rebellion, seem to me well taken, and good illustrations of the negative principle. The maxims given in (a) and (c) are patent sophistries attempting to rationalize plainly unjust courses of action. Preventive war (b) is generally recognized as a violation of international right.⁸

4.9 Rightful Freedom and Beneficial Consequences

It is important to see that Kant is not claiming that those policies that pass the positive and negative tests of publicity are necessarily the wisest or the best policies, or those most conducive to people's welfare or happiness, or the policies a wise politician would most like to see adopted. Politicians will be forbidden by Kantian principles of right to achieve the best consequences when these come at the cost of the union of ends on which the state must rest. Suppose that the total wealth of society can be maximized only if it is distributed in such a way that some are destitute, lacking the means to live, or lacking the property necessary for them to live freely and independently (TP 8:295). These people cannot then rationally consent to this distribution, which would violate the idea of the original contract that provides the criterion of justice (TP 8:296–7, 304–5; MS 6:340–1). Politicians will in that case be required by justice to tax the rich to support the poor (MS 6:326), even

⁸ The 2003 U.S. invasion of Iraq has been defended as a case of justified preventive war by representatives of the Bush administration, who argue that the prohibition on preventive war is unrealistic. However, in this case the Bush regime apparently did not conceal its intentions, since in the run-up to the war, there were many large popular demonstrations, both in the U.S. and internationally, protesting the planned invasion of Iraq and urging the Bush government not to take this unwise and illegal step. In any case, it is plainly false that this was even a case of preventive war. For in fact Iraq posed no threat whatever to the U.S.: the U.S. claims about Iraq's actual or imminent possession of "weapons of mass destruction" (ostensibly aimed at the U.S.), and the so-called "intelligence" cited to back up these claims, were transparent lies, or at best cases of gross self-deception. The real purpose was a war of conquest aimed at taking possession of Iraqi oil reserves in the interest of Western corporations (the regime's political backers). The defenses of the invasion of Iraq based on doctrines of preventive war were therefore doubly sophistical: first, the international legal principle forbidding preventive war is quite valid, for precisely the reasons cited in Kant's argument about example b); and second, the invasion of Iraq was not a preventive war at all, but a case of patently illegal aggression and conquest, like example c).

if this means that the wealth of society as a whole is less than it might otherwise have been. This redistribution of wealth, moreover, would be a matter of right, not of charity or beneficence, since duties of beneficence are ethical duties only, and not subject to coercive enforcement (EF 8:385; MS 6:454).

It might happen that the policies that would have the best consequences for all would meet with resistance from the population, owing to their foolishness or misguided ideas. Kantian moral politicians would then try to persuade them of the wisdom of these policies; if they cannot, then the policies with the best consequences could be successfully implemented only secretly—for instance, by Dostoevsky's benevolent Grand Inquisitor, or by an elite coterie of utilitarian social engineers who would maximize the general happiness, if necessary, even against the will of those they would benefit. In that case, however, these policies would violate the negative principle of publicity, and Kant holds that politicians should not try to implement them.

The first task of political rulers, in Kant's view—indeed, their only fundamental task—is to maintain the condition of right or justice, that condition in which each retains that freedom (independence of the will of others) that is compatible with a like freedom for all according to universal law. In Kant's view, justice does often lead to happiness, and the best way to achieve justice is also often to make people happy, at least when there obtains that fortunate state of affairs in which what people think will make them happy will also in fact do so. But Kant regards a paternalistic government—one that decides for its subjects what their happiness is to consist in, and then imposes it on them against their will (either by force or by deception)—as “the greatest despotism thinkable” (TP 8:291). Politicians who insist on producing the best consequences by any means available, even by secrecy, deception, or the violation of the rights of citizens, could do no wrong if they were governing only beings, such as domesticated animals, who are incapable of managing their own lives by free choice—beings with neither rights nor dignity. They should never be given political authority over human beings.

At this point, anyone who argues as I just have will no doubt be presented with alleged historical examples where the constraints of justice might seem to have required a politician to forego a course of action that was allegedly the only way of achieving some important good or preventing some disaster. The conclusion will be that Kant's principles are obviously too constraining.⁹ In assessing such

⁹ The example suggested to me here by William Scheuermann is Franklin Roosevelt's declaration in 1940 that he would never send Americans to die in the war overseas. This appears to have been a lie, but perhaps a necessary one to get the U.S. population eventually to accept the inevitability of war as the only way to defeat the Axis powers. One might allege that without such lies, the Nazis might have been victorious in Europe. If Kant's principles would require that we allow such a calamity as that, it might be claimed, then they are obviously and disastrously wrong. One possible reply to this example,

objections, it is crucial to appreciate that this sort of hindsight badly distorts our moral vision, and in more than one way.

First, there are obviously many wrongful acts whose results, in historical retrospect, we are glad happened. We must not change our view about the wrongfulness of the act just because its consequences are welcome. Rather, we should feel deeply ambivalent about this act, continuing to condemn it, while also refusing to regret its good consequences. Kant's entire philosophy of history is based on the idea that human competitiveness—unsociable sociability, the radical propensity to evil—is nature's device for developing the faculties of our species—including the very moral reason that enables us to recognize and struggle against this propensity (I 8:18–22). Thus it is deeply Kantian to acknowledge this kind of ambivalence. Second, looking at matters retrospectively can often make it seem that a wrongful action was the only possible way of achieving some progress or averting some calamity. But we can never know the truth of all the possible counterfactuals in such a case. Rightful ways of achieving the same result might have been found or invented by a more sagacious politician. Third, still less can we claim to know this about some decision that has yet to be made, whose success or failure, as well as its unforeseeable good or bad results, still lie in the contingent future.

Kant does think, as we have seen, that politicians should aim at good results. From the standpoint of political prudence, he thinks they should always look at the consequences of their policies in the spirit of a Weberian "ethics of responsibility." He merely insists that the claims of right are prior, because conformity to them is the condition for the legitimate possession of whatever powers they have to bring about their ends. Kant's solution for reconciling politics with morals therefore maintains that it is always the politician's task to find some way of achieving the good result (or averting the bad one) that remains consistent with the demands of right. It can never be known in advance that this is impossible.¹⁰ Politicians may protest that Kantian principles of right make their task more difficult, cutting off the easiest and most certain route to their (noble) ends. But right is not supposed to make the politician's life easy; it is rather to make possible a condition of freedom for all under universal laws, which is the sole ground of any legitimate powers politicians ever possess. The easiest life for a politician would be the life of a

however, is that Roosevelt might not have been lying in 1940, since it was still unknown whether U.S. entry into the war was avoidable. He may have been doing his best to report his hopes and intentions truthfully, and it is only in retrospect—in light of later events that he could not have known at the time—that we can think he was obviously lying. This is a third way—besides the two I am about to discuss in the text—in which appeals to hindsight can involve a profound distortion of our moral vision.

¹⁰ I am grateful to Sandra Shapshay for suggesting this reply to the present objection, which is obviously the only authentic Kantian one, and also (I believe) entirely convincing.

tyrant untrammelled by right. The farther we move away from that, the harder it may become for politicians to achieve their ends, whether these are good or evil.

The shabbiest and most pitiful rationalization politicians can ever offer for their wrongdoing is the wishful speculation that although most people now condemn their actions, someday these actions will be “vindicated by the verdict of history.”¹¹ Of course, it can never be known either way whether future people will make such a judgment. Even if they do, this would not justify wrongdoing. In many matters, history’s verdict may be unjust, condemning rightful actions and praising wrongful ones. The verdict of one future age may always be overturned (for better or worse) by later revisionist historians or by common opinion. But all of this is beside the point, because no action that is truly wrongful could ever be justified merely by its historical outcome. The only way any “verdict of history” could ever truly justify such an action would be if future people correctly decide, on the basis of reasons having nothing to do with the outcome of this particular action, that the moral principle it violated was a mistaken principle.

4.10 Kant on Lying—and Political Lying

The real philosophical issue raised by Kant’s principles of publicity is whether it is indeed a requirement of justice that statesmen and politicians should make their policies public and speak the truth about them. This requirement certainly coheres with the spirit of a Kantian theory of right, which is based on the claim that standards of right serve to protect external freedom (the independence of every person’s will from constraint by another’s choice) as far as this is possible according to universal law. For the falsification and concealment of someone’s intentions, when this prevents another from exercising his own choices freely and with necessary information, is an obvious way in which the choice of the former effectively constrains the freedom of the latter. If both the subjects of a commonwealth and the other states who make treaties and agreements with it are not safe from such falsification and concealment, their freedom is unjustly infringed.

Kant’s principles of publicity are continuous with (though an extension of) his assertions in the (in)famous (because infamously misunderstood) essay on the right to lie, that the public assertions of politicians are *declarations*, in the sense bearing on right. This is not the place to give a full discussion of the issue (which I have done elsewhere: Wood, 2008, Chapter 14, especially pp. 240–6, 249–51). But

¹¹ This kind of rationalization is so common and so disgusting that it deserves a special name. Call it “self-congratulation in the future perfect tense.” It is this lame apology that, for instance, G. W. Bush and his admirers now offer for his evil and disastrous decision to invade Iraq in 2003.

the following summary of that discussion seems both necessary and sufficient for present purposes.

Kant's assertions about the ethical duty to oneself not to lie (which seem, at least at first glance, to be extremely strict) (MS 6: 429–31) are thus simply not relevant to political contexts—or, indeed, to the context of the (in)famous right to lie essay. To confuse the two is to make precisely the mistake made by the *Realpolitiker* who think that Kant's views about the duties of politicians are drawn from “private morality” or a *Gesinnungsethik*.

Kant's position on truthfulness in matters of right (which is all that is relevant to the “morals” applicable to politics) is that it is forbidden to make false declarations (VRL 8:426). “Declaration” (*Aussage, Deklaration*, Lat. *declaratio*) is a technical term in Kant's theory of right, referring to any statement on whose truthfulness others are entitled (*befugt*), as a matter of right, to rely (cf. KpV 5:44; MS 6:254, 258, 272, 304, 366). To speak untruthfully, in such contexts, is to deprive others of something to which they have a right, something that takes away their rightful freedom and subjects them unjustly to the liar's choice in matters that by right belong to their free choice. Paradigm cases of a declaration might be a statement under oath in court, or a warranty included in a contract. To give false testimony when your word is supposed to be taken as probative in court is to deprive judges, juries, and the entire system of what they need by right to do their work (see KpV 5:44). To falsify claims stated in a contract, or to contract to meet conditions you know you will not meet, is to deprive those with whom you contract of their rightful free choice in deciding to deal with you or not deal with you on the terms stated (MS 6:272).

A lie (*Lüge, mendacium*) for Kant is a technical term, meaning an intentionally false statement that is a violation of a duty of right. There is, for Kant, no violation of right involved in falsifying statements that are not declarations, statements made in contexts where no one has a right to rely on their truthfulness. This occurs when one “is merely communicating his thoughts to them, telling or promising them something, whether what he says is true and sincere or untrue and insincere (*veriloquium aut falsiloquium*); for it is entirely up to him whether they want to believe him or not” (MS 6:238). To say that you will give someone something, for instance, with no consideration on the other side and no anticipated loss to the other person if you do not give it, is not to make a declaration. No violation of right occurs if you fail to keep such a promise. The context of human dealings (where necessary for the sake of clarity, the terms of positive legislation) must determine whether or not what someone says counts as a declaration.

The public statements of politicians, their announcements of their policies, and the factual statements they offer in support of them, are regarded by Kant as declarations. For politicians to speak falsely in these contexts, about matters of

public right, is therefore contrary to right. That is the real issue between Kant and Benjamin Constant in the controversy about the right to lie. This is clearly indicated by the title of Constant's essay to which Kant is replying: *Des réactions politiques* (Constant, 1964, pp. 63–71; cf. VRL 8: 429). In contrast to Kant—but in line with the assumption our *Realpolitiker* usually make about the standards moralists want to apply to politics—Constant thinks that these standards are derived from private morality. Constant's position, contrary to the realists, is that moral duties can be applied to politics, but (he thinks) only conditionally. His example is that the duty to tell the truth applies to politics only under the condition that those to whom one speaks have a "right to the truth" (VRL 8:425).

Kant's opposed position is that, first, the duties of politicians as politicians are duties of right, which are entirely separate from duties of ethics (or "private morality"). And in the case of the duty to be truthful, the duty of right attaches not to statements in general but only to declarations, and further to any declaration as such. This duty, moreover, persists even if the person to whom you make it has no right to the truth, so that there is no assignable obligation *to that person* to tell the truth. For instance, a person might lack such a right because their own unjust behavior or intent has forfeited it. Even in such a case, however, Kant holds that you owe it to the right of humanity generally not to falsify declarations. This is how Kant applies it in the case of the man who appears at your door with murderous intent. This unjust intent might deprive *him* of any right to your truthfulness, but if your statement to him is made in a context that makes it a declaration, then you do wrong to humanity by falsifying your declaration. That your statement here is assumed to be a *declaration* is a highly significant assumption that Kant is making about this example, for the sake of argument, in order to bring out the issue between himself and Constant. That Kant is making this (arguably very artificial and implausible) assumption is almost universally overlooked in discussions of Kant's position on the right to lie, which usually react to the example taking for granted quite different (doubtless more natural) assumptions about it. The point on which Kant means to focus is that the public, humanity generally, has a right to rely on the truthfulness of declarations, since they concern matters of people's rights.¹²

¹² Kant's claim here about declarations made to a person with no assignable "right to the truth" is that you do not wrong him by falsifying your declaration but you do violate a principle of right. This is analogous to the claim, which we saw Kant making earlier, that deposing a tyrant by force might not violate his right, but it is nevertheless a wrong or injustice. Kant's general principle is that the fact that someone has done an injustice (even an injustice to you) does not make it right for you to do the same injustice to them. To permit this, he argues, would involve the nullification of all principles of justice: "When one country has broken the peace, the other cannot do so in retaliation, for if that were

The crucial claim here is that the public statements of politicians and statesmen are *declarations*. This is because people's rightful political choices (what policies or politicians they support, for example) must depend on truthful information about what politicians believe, the principles they stand for, the policies they favor, and the like. To expect those who discuss politics and politicians in a public forum, or who vote for or against them in an election, to make such choices based on the statements of politicians, while not regarding these statements as declarations, is like expecting judges and juries to treat testimony as probative while regarding it as permissible for witnesses to lie at their discretion.

In the Appendix to *Toward Perpetual Peace*, Kant may be seen as extending further the basic idea that political speech involves declarations, to the idea that publicity in general is a condition of political right. Kant's principles of publicity extend the requirement, claiming not only that politicians behave unjustly if they make false statements, but claiming also that their policies must be such that they could be publicly declared without thereby becoming self-defeating, and further taking as a sufficient condition of the rightfulness or justice of their policies that they would require publicity in order to be effective. Kant does not say that politicians must make their policies public—only that the policies themselves must be *capable of publicity* on pain of anticipated self-defeat.

4.11 Right and Practicality in Politics

No doubt it would be naïve to expect either politicians or witnesses in court always to live up to their duties of right regarding truthfulness. Kant is not the least bit naïve about this. His point is rather that if we do not acknowledge the presence of genuine duties of right in both cases, then we no longer have a system of right at all, but only a sham or pretense of one. We fail in practice to acknowledge these duties, thus creating such a sham, if we treat the habitual violation of these duties by politicians or witnesses only casually and with indulgence, as though the duties they are violating were not real duties at all (so that the desideratum of truthful speech is perhaps only one consideration among others that might rightfully be overridden

allowable, then no peace would be secure. And though [a lie to a deceptive or wrongfully malicious person] may not infringe [his] particular right, it is still already a lie [in the sense of a false declaration], and contrary to the right of humanity" (VE *Vigilantius* 27:447). Another possible application of the same idea (though not one Kant specifically discusses) might be to the issue of torture. If torture represents a wrongful kind of action (a violation of the right of humanity, say) then it will still be wrong, even if the person tortured has deprived himself of the assignable right not to be tortured, as for instance by being himself a torturer, or a terrorist, or some other perpetrator of egregious injustice. What matters in that case, on Kantian principles, is not the specific rights of the person tortured, but the wrongfulness of the act of torture itself.

for the sake of witness's or the politician's personal moral sense or political ends). Such a sham may, of course, all too often be what we do in fact have in the real world. But we also accord political and judicial processes a kind of authority that would be incompatible with regarding them as nothing but arenas of coercion and manipulation unregulated by right. Kant also very plausibly supposes that in the long run no stable republican political order can exist if political institutions are generally regarded in the latter way.

There is no coherent "compromise" at this point. If the actions of politicians are not limited by right, then, if we are consistent, we must deprive all politicians of their authority to do business in the public's name (which would render a condition of right impossible). We merely contradict ourselves if we accept as valid the assumption that we are operating in a system of right, but then treat the duties constitutive of this system as subject to justifiable exceptions on independent grounds, at the discretion of politicians, merely on the condition that they find they can get away with violating these duties with impunity. Our *Realpolitiker* complain that moralists would impose their private values on politics; what they impose is far worse: arrogant caprice masquerading as "responsibility." Weber portrays this as the strong-willed decisiveness of a mature man who rationalizes his abuse of power with the self-important thought that he is shouldering the heavy burden of responsibility: "It is immensely moving when a mature man . . . is aware of a responsibility for the consequences of his conduct and really feels such responsibility with his heart and soul" (Weber, 1958, p. 127). In real life, however, such people always end up sounding like the rogue Marine commander Colonel Nathan Jessep (played by Jack Nicholson) in the 1992 movie *A Few Good Men*:

You want the truth? You can't *handle* the *truth*! I have a greater responsibility than you can possibly fathom. You have the luxury of not knowing what I know. You don't want the truth, because deep down in places you don't talk about at parties, you *want* me on that wall, you *need* me on that wall. We use words like "honor," "code," "loyalty." We use these words as the backbone of a life spent defending something. You use them as a *punchline*! I have neither the time nor the inclination to explain myself to a man who rises and sleeps under the blanket of the very freedom that I provide, and then questions the manner in which I provide it! I would rather you just said "Thank you," and went on your way. Otherwise, I suggest you pick up a weapon, and stand a post. Either way, I don't give a *damn* what you think you are entitled to!

Dress it up any way you like: this is still nothing but the arrogant, wrongful abuse of power. In the mouths of these criminals, words like "honor" and "responsibility" are worse than a "punchline"—they are obscenities. We do not *need*, and should not *want* such people—not on *their* terms: on those terms, they are always a *menace* to our freedom. And from them we are always *entitled* to the *truth*: whether we can "handle" it is not their call.

The point here is not that there is a set of “absolute” rules, that must never be violated, or that there are not sometimes tragic dilemmas, in which whatever the agent does is wrong, or that it is never necessary to do repugnant actions. The point is rather that people who are too eager to believe that all these cases occur, and are occurring in their lives, and that they are the ones who must bear the burden of them—those are precisely the people who should never hold political power, or have the rights of others exposed to them. Jessep says: “You have the luxury of not knowing what I know.” What Jessep claims to “know” is that terrible actions are necessary, that those who insist on human rights are too squeamish, they are “weaklings”—that we must descend to the level of evil people in order to defeat them. This “knowledge” makes them feel strong, superior to the ordinary mortals over whose fate they exercise their power, always, of course, with the proud awareness of the “responsibility” that goes with it.

“A too great readiness to think ‘I can’t do anything but this terrible thing, nothing else is open to me’ is a mark of vice, a flawed character” (Hursthouse, 1999, p. 87). It is a vice that absolutely disqualifies anyone afflicted by it from exercising political authority over others. Clearly, it disqualifies Jessep, and all his real-world counterparts. To them, the most terrible and wrongful actions—torture, targeted assassination, indefinite detentions without charge or trial—are not mere exceptions to be allowed in extreme circumstances, but become matters of regular national policy, so that in the end it is only their jingoism, or their racial and ethnic bigotry, that enables them to distinguish themselves, even in their own minds, from the “totalitarians” and “terrorists” who have lured them into disgracing their humanity. Would that their careers could all end ignominiously, with the final line spoken to Jessep by the JAG lawyer Lt Daniel Kaffee (played by Tom Cruise): “You’re under arrest, you son of a bitch.”¹³

¹³ The events depicted in *A Few Good Men* are based on a real military court case that occurred in 1986. The real-life counterpart of Lt Kaffee was David Iglesias, later a U.S. Attorney for the District of New Mexico, who was one of eight U.S. attorneys summarily dismissed in 2006 when they refused to prosecute meritless cases of “voter fraud” as part of a Bush administration plan to suppress voting among young, poor, and minority voters, so that the Republican party might maintain electoral superiority, even though its policies violate the rights and run directly contrary to the interests of the vast majority of Americans. Other such efforts are ongoing. They often perpetuate, mimic, or renovate analogous practices by the Democratic party in southern states of the U.S. prior to the Civil Rights Era. There are also efforts underway to doctor the long-outdated American electoral college system—but only in certain states—in order to make more likely the election of a president for whom only a minority of the electorate has voted. These efforts are evidently occasioned by the fact that in the last presidential election, the majority that voted for the victor included less than 40% of the white population, many of whom apparently regard themselves as constituting exclusively the only true legitimate American electorate.

4.12 The Moralism Speaks to Practical Politicians

Kant wrote *Toward Perpetual Peace* shortly after being reproved by education minister J. C. Wöllner, on behalf of King Friedrich Wilhelm II, for his writings and teachings on religion. In this work, he went out of his way not only to praise the constitutional form of the newly created French Republic, which until recently had been at war with Prussia (and was soon to be at war with it again), but also to suggest, impudently, that a nation that had raised its constitution to the form of a republic might serve as a suitable rallying point for other nations in creating a federation with the aim of preserving a permanent and just peace in the international sphere. Such opinions were about as popular with the rulers of Prussia as advocacy of health care reform, progressive taxation of the rich, or indictment for fraud of Wall Street bankers would be if these were presented today before the Republican leadership in the U.S. Congress.

Kant had no illusions on this score either. He knew he was admonishing and offending those who did not care a fig for his opinion, and he did not for a minute expect that they would pay any heed to what he was saying. His real intended audience had to be those both within Prussia and outside it who would overhear his moralistic complaints against the Prussian powers that be, and might share his disapproval of them. Kant did not believe in revolution, or even in civil disobedience; a year earlier he had acceded to the King's command not to teach or write on the subject of religion, promising to obey it. He wanted to be seen as an obedient subject of the Prussian King. But he also did not believe in holding his tongue when he could get away with speaking truth to power. He made his promise of obedience only after offering a defense of his position, and then less than two years later he resumed writing on religion after the death of Friedrich Wilhelm II had released him (as he saw it) both from the royal command and from his own promise. As Paul Formosa has pointed out, Kant denied the right of positive or active (forcible) resistance to legitimate but unjust authority, but he affirmed, both in principle and in deed, the right and even the obligation to offer "negative resistance," by way of public protest (Formosa, 2008). He even considered a people "corrupt" if it accedes blindly "to every demand the government puts forth" (MS 6:322).

Kant even seemed to relish his role as the "impractical" theoretician or moral philosopher impudently (though perhaps also impotently) admonishing practical politicians concerning their duties. He always hoped on principle that those in power would listen to him, but never really expected them to (see EF 8:368–9). It was no doubt easier to maintain such an attitude in a state that entertained no pretense that rulers would ever need to listen to reason, respect the law, or pay any attention to the views of their subjects. Despite such pretenses, it is still true in our

day that gross abusers of political power seldom have anything to fear from the law or the public—unless, perhaps, they were to appear hostile to the interests of great wealth or corporate power—and then it would be this, even if done in direct pursuit of justice, legality, and the public interest, that would put politicians in danger. What is hardest to accept is only that naked injustice still rules so much of our politics that our situation remains very much the same as Kant's.

Herder and Kant on History: Their Enlightenment Faith

5.1 Herder and the Enlightenment

One of Kant's least attractive traits as a human being was a tendency to regard his students, followers, and protégés as disloyal to him if they departed from what he saw as the central tenets of the critical philosophy. Kant displayed this ugly trait especially towards two men—Johann Gottlieb Fichte and Johann Gottfried Herder—through whom Kantian philosophy was greatly enriched and its influence extended far more than it would otherwise have been. Both were great and original philosophers in their own right, but also touchy, difficult, impossible personalities, far more flawed in dealing with others than Kant ever was. Fichte was a better friend to the critical philosophy than Kant ever realized, however, often drawing conclusions from Kantian principles far more consequentially than Kant himself did. And Herder's thought, along with Kant's, helped to revolutionize the study of human nature and history. In both cases, the personal conflicts have led people to imagine deeper philosophical divisions than I think are really there.

Herder was Kant's student, and most of what we know about Kant's lectures on metaphysics and ethics between 1762 and 1764 comes from lecture transcriptions in his hand. Herder came to prominence when his essay *On the Origin of Language* won a prize from the Prussian Academy in Berlin in 1771, but his most important contributions to the philosophy of history were the short essay *This Too a Philosophy of History for the Formation of Humanity* (1774), and the massive *Ideas for the Philosophy of the History of Humanity* (1784–1791). Herder's contributions to anthropology and philosophy of history thus predate Kant's, and the mutual influence between the two thinkers is clearly greater than is commonly appreciated. Kant's personal relations with Herder were complex. Kant's reviews of Herder's *Ideas* were superficially laudatory, but insultingly condescending. Herder's last

works, *Metacritique of the Critique of Pure Reason* (1799) and *Kalligone* (1800), were polemics against Kant. Yet in the late 1790s, Herder's *Letters on the Advancement of Humanity* included an eloquent tribute to Kant, which Lewis White Beck used to quote at the conclusion of the prefatory "Life and Works" section in his translations of Kant.¹

Many treatments of Herder see him mainly as a follower of Kant's eccentric counter-enlightenment friend Hamann, and an enemy of Kantian philosophy. Herder is characterized as a figure of the *Sturm und Drang*, a counter-enlightenment thinker, a nationalist rather than a cosmopolitan, a Francophobe and Germanophile, philosopher of feeling and intuition rather than reason, a cultural relativist and an opponent of moral universalism. But these images represent half-truths at most. Fred Beiser seems to me to get closer to the truth when he describes Herder's philosophy as resulting from a struggle between the twin influences of Hamann and Kant, and also when he concludes that if either of Herder's teachers could be said to have won the struggle, then it was Kant (Beiser, 1992, Chapter 8).

No doubt Herder and Kant disagreed about a number of things in philosophy. Michael Forster puts forward the interesting thesis that Herder derived his main philosophical opinions from Kant—but the pre-critical Kant of the 1760s, not the critical Kant of 1781 and after, with whom he had some profound differences. As Forster sees it (PW pp. xi–xiv), Herder accepts Kant's early (Hutchesonian) moral sense or (as Forster anachronistically puts it, "nonscognitivist") views in ethics, his "(Pyrrhonist-influenced) skepticism about metaphysics" and a brand of empiricism that Forster supposes the pre-critical Kant to have held—all of which positions are starkly opposed to the Kant of the later critical philosophy. Here I would sooner question some of Forster's characterizations of the early Kant than his attribution of these positions to Herder. But on the philosophy of history, which is my present theme, I think we can properly appreciate the real differences only if we begin with the recognition of how much Kant and Herder have in common, and how even their quarrels in the philosophy of history rest on more basic agreement.

The main point on which Kant and Herder agree, I will argue, is that it is rational to look at human history as exhibiting a kind of natural purposiveness, like that found in organisms, rather than that found in intentional human actions. Both philosophers, I will argue, should be seen as grounding their philosophy of history

¹ See Beck's introductions to: *Prolegomena to Any Future Metaphysics* (Indianapolis: Bobbs-Merrill, 1950), p. xxii; *Critique of Practical Reason* (Indianapolis: Bobbs-Merrill, 1958), p. xxii; *Foundations of the Metaphysics of Morals* (Indianapolis: Bobbs-Merrill, 1959), p. xxii; *Kant on History* (Indianapolis: Bobbs-Merrill, 1963), p. xxviii. The quotation is from Herder, *Briefe zur Beförderung der Menschheit* (1793–1797), *Werke* Suphan 18:324–5.

on what could be described as a faith in human progress. This is a faith that is characteristic of (though by no means universal in) the Enlightenment. Sometimes Enlightenment faith takes a religious form, sometimes a purely secular form, and sometimes a mixture of the two (as in both Kant and Herder).

In our time, it seems to be fashionable to criticize or condescend to Enlightenment views of historical progress, as though they represented some sort of naïveté that we—who have lived to see such twentieth-century events as the Nazi Holocaust, the bitter fruits of European imperialism, and the collapse of socialism—can now recognize as a kind of illusion. Postmodernism now dismisses such Enlightenment views as “meta-narratives,” and many religiously inspired views have come to see them as exhibiting a kind of secular humanist hubris. For this reason, after presenting the views of Herder and Kant, I will conclude with some reflections on their faith in historical progress. To this end, I will later call upon some of Robert Adams’ thoughts about *faith*—especially the kind of “moral faith” he discusses in the last chapter of *Finite and Infinite Goods* (1999). I will try to use Adams’ account to make the case that the Herderian and Kantian views of history are neither objectionably naïve nor hubristic, but quite reasonable. Enlightenment faith in historical progress, I will argue, is a kind of rational faith (in Adams’ sense of “faith”) that is rationally necessary for us as often as we seek a rational understanding of human history and of our own actions and strivings as part of it.

5.2 Herder’s Historical Manifesto

Kant is well known for the range of his philosophical interests and accomplishments. His philosophy revolutionized virtually every field in philosophy: metaphysics, epistemology, philosophy of mind, philosophy of science, moral and political philosophy, aesthetics, philosophy of religion, and (as I am arguing here) philosophy of history. It is much less well known that very much the same could be said of Herder in many fields of the human studies, such as philosophy of mind, philosophy of language, philosophy of action, the methodology of history and anthropology, and both biblical and literary interpretation. The theoretical revolutions of Schleiermacher in hermeneutics and Wilhelm von Humboldt in linguistics were built directly on Herder’s achievements (see Forster, 2007).

In 1774, Herder published *Auch eine Philosophie der Geschichte zur Bildung der Menschheit*—an emotional and declamatory little essay, wildly ambitious in scope, rich and original in content, by turns impassioned and sarcastic in tone. The standard reading takes it to be an anti-Enlightenment polemic. Herder’s essay is filled with indictments of his own age, regarded as an age of Enlightenment, an age of philosophy. It is an age of abstract intellect, in which the head is divided against the

heart, in which a superficial, artificial, mechanical way of thinking has cut people off from their deeper humanity. It is an age of skepticism, of religious unbelief, of abstract cosmopolitanism, in which the values of patriotism and cultural rootedness have been lost. Finally, this is an age of narrowminded complacency, which looks down from the supposed height of its philosophical wisdom on all earlier ages, self-conceitedly regarding itself as the final goal of human history. All these traits are symptomatic, Herder seems to be saying, of an exhaustion of the powers of life.

By contrast, Herder praises the virtues of past times, especially those to which he thinks his age condemns and to which it feels superior—the patriarchal religion of the Orient, the priestcraft of Egypt, the patriotism of Greece and Rome, the dominance of religion over all spheres of life in the long era of Christianity that preceded the Renaissance and Reformation. Much of this polemic against his own time is continuous with the works of literary criticism that preceded his historical manifesto. In them, Herder argued that ancient languages, such as that of Homer and the Bible, were more poetic than modern languages, which had become too intellectualized, too rule-governed, too cut off from everyday life. He celebrated the variety of folk-culture over the classicism that saw beauty only in the imitation of a narrow range of privileged models taken from Greco-Roman culture.

Above all, Herder defended German culture against domination by the French—a feature of his work that later led the Nazis to claim him, along with Fichte, as one of their own. In both cases, of course, such a claim is false to the point of obscenity, but it is especially distorting in the case of Herder, whose supposed “nationalism” involved an intense hostility to most of the things—such as state and military power—that that term connotes. Besides, the proper calling of German culture in Herder’s view was always thought, poetry and teaching, and for him “German” culture included even Shakespeare and the supposed Celtic folk poetry of Ossian (which Herder regarded as authentic, even after cooler heads, such as Hume’s, had immediately recognized MacPherson’s crude forgeries for what they were). For another thing, Herder’s Germanophilia was always cultural, never political, and least of all military.

Herder’s historical manifesto of 1774 is often read exclusively as a tirade against the age of Enlightenment, partly because there are many who find the position it puts forward, so interpreted, highly attractive. According to them, Herder sees through the arid intellectualism of us philosophers, recognizes the shallow elitism that hides behind the Enlightenment’s façade of egalitarianism. He is (they think) in touch with our deeper humanity, which is denied by the shallow scientific rationalism of the eighteenth century, and so on and so forth. Herder is then seen as the harbinger of the Romantic movement, the earliest perceiver of the

truths articulated only recently by multiculturalism and post-modernism. Many of Herder's admirers find all this appealing, or at least some of it.

I would not want to deny that Herder's thought does anticipate much in these later movements, or that it has exercised much influence on them (much of it indirect and even now insufficiently acknowledged). But this reading of Herder still seems to me profoundly wrong, not only in its view of Herder, but even more in its orientation to modernity. To begin with, it leaves people like me far too easy a way out—I mean, those of us who embrace the Enlightenment precisely *because* it was an age of skepticism and unbelief, *because* it rejected patriotism and sentimentalism, *because* we think the intellectual honesty of natural science exhibits more vitality than the loathsome emotionalism of preachers and poets—the latter accurately pegged by Hume as “liars by profession” (Hume, 1967, p. 161). To read Herder simply as “counter-enlightenment” merely reduces the issues between us and him merely to a matter of taste: by which I mean bad taste on his part, good taste on ours.

Then too, on this reading we can reject a lot of what Herder seems to be saying about the Enlightenment as simply a gross mischaracterization. Which Enlightenment thinkers are supposed to be self-complacent, thinking that the whole of history leads up to them as its pinnacle and final end? Some of the Enlightenment's leading figures, such as Voltaire and Mendelssohn, don't seem to believe in historical progress at all. Those who, like Kant, do cautiously entertain the idea of historical progress—more cautiously, in fact, than Herder did—focus their hopes on the future, not on the historical present—exactly as Herder himself does in the closing pages of *This Too a Philosophy of History*.

As this last point makes clear, the exclusively counter-Enlightenment reading also makes it all too easy to cite Herder against himself, since he himself is a partisan of Enlightenment values just about everywhere it counts. Herder's religious faith turns out to be that of a typical German freethinker of the time—highly personal and heterodox, laced with Spinozistic heresies and idiosyncratic, non-literal readings of the Bible. Herder is an advocate of toleration and open discussion in all matters, anti-authoritarian, anti-militaristic, anti-imperialistic, egalitarian in his social, economic, and political convictions. Herder's chief aim in the human sciences was to take better account of the empirical facts, to expand and adjust scientific method, not to reject it.

What we need, therefore, is a reading of Herder that explains why he polemicizes so ardently against some prominent representatives of the Enlightenment, such as Voltaire and Kant, and shows us what he wants to add to the Enlightenment, while acknowledging the equally plain fact that when the chips are down, he is himself a part of the Enlightenment.

5.3 Understanding Each Culture from Within

The best way to reach a better reading of Herder's 1774 historical manifesto is to begin with its most revolutionary idea. This is Herder's new way of looking at cultures and ages different from our own. Every age of human history, Herder claims, must be understood from within, in terms of its own "way of thinking" (*Denkungsart*), which includes its own ways of experiencing and feeling, as well as of living and acting. Each age is unique, to be understood first of all simply in terms of itself and its inner organic forces, like a living thing or a work of art, and not mechanistically, as simply an instance of some general laws. As he puts it later in the *Ideas*: "The feelings and inclinations of human beings are everywhere conformable to their organization and the circumstances in which they live; but they are everywhere swayed by custom and opinion" (Herder Suphan 13:319, *Ideas*, Book VIII, Ch 4).

In the first section of *This Too a Philosophy of History*, Herder reviews the main cultures of the ancient world and attempts to characterize each in its unique historical place. He distinguishes them in terms of their economic mode of life, but also in terms of the stage of development of the human being that they represent. "The orient"—that is, the ancient near east—was a tranquil pastoral society, grounded on submission to patriarchal authority and childlike religious feeling (Herder Suphan 5:477–80; PW 273–76). Ancient Egypt was an agricultural society, which developed landed property, civil administration, and the practical arts (Herder Suphan 5:489–92; PW 280–4). The Phoenicians, in contrast to the Egyptians, were oriented outward, towards the sea and towards trade between peoples (Herder Suphan 5:493–4; PW 284–5). The achievements of both were taken up, but also entirely transformed, by the Greeks, in whom there first arose a new sense of beauty and a love of freedom (Herder Suphan 5:494–9; PW 285–9). With the Romans, we see a pride in conquering one's desires and sensual gratification, devotion to the fatherland, and the aspiration to build an entire world (Herder Suphan 5:499–501; PW 289–91). Herder also compares the successive ancient cultures to stages in the development of a single human being. The Orient is the infancy of the species, Egypt its boyhood, Greece its youth, Rome its coming to manhood. (In these terms, he significantly depicts his own century as an old man, a point to which we will return later.)

Each way of thinking represents a distinct side of humanity, and seeks a kind of happiness distinct from the others (Herder Suphan 5:508–9; PW 296). For this reason, Herder observes that every people regards the way of thinking of all others with intolerance. The agricultural Egyptian despises both the Oriental shepherd and the Phoenician seafarer; the Greek regards all other peoples as barbarians

(Herder Suphan 5:487; PW 281). Yet “prejudice is good in its time, since it renders happy. It forces peoples together into their center, makes them firmer on their tribal stem, more blooming in their kind” (Herder Suphan 5:510; PW 297).

There is one serious danger here, of which Herder is aware, and which he takes steps to avoid. We might be tempted to ascribe to him the view that ways of thinking and ways of being happy are essentially cultural and collective, rather than individual. But this is the exact opposite of what he holds. “No one in the world feels the weakness of general characterizing more than I,” he declares (Herder Suphan 5:501; PW 291). All thinking, and feeling, and the constitution of any human happiness, is always at bottom an expression of the distinctive individuality of particular human beings. In some ages there is greater uniformity, greater dependency on social authority or conformity to collective standards than in others (more in Oriental or Egyptian than in Greek culture, and more in all of these than in modern culture). But in every culture, Herder holds, there is more individual variation in thinking and feeling than there is cultural uniformity. This is most true of all when it comes to what is most fundamental to human life, the conditions for happiness. As he says later in the *Ideas*: “Happiness is an internal state; and therefore its standard is not seated without us, but in the breast of every individual, where alone it can be determined” (Herder Suphan, 13:333; *Ideas*, Book VIII, Ch. 5).

Herder’s interest in generalizing about ages and peoples, and distinguishing different cultural ways of thinking and feeling, revolves around two main points. The first is the radical differences between their ways of thinking and consequently of the conditions under which life—and a flourishing and happy life—is possible for each. These give rise to contrasting conceptions of morality and virtue that differ from one another just as radically as the mode of life and the happiness possible to each. The second point—which Herder refers to at one point as “my great theme”—is the historical continuity between the different ages, the necessary development from each stage of history to the next, which, Herder claims, displays the plan of Providence (Herder Suphan 5:511; PW 298–9).

5.4 The Failings of the Present Age

Herder views his own age as approaching its own history in an entirely wrong way, because it applies its own way of thinking, its own values, its own conceptions of happiness and virtue—in a word, what he calls (using a typical Enlightenment term) its own “prejudices”—to other ages—for which they provide the wrong measure and result only in uncomprehending hostility. The present age, for instance, identifies the so-called “despotism” of the ancient Orient or the hierarchy (priestcraft) of the Egyptians with the closest approximations to them in our

own time, and because it rightly rejects these in relation to our happiness and our way of thinking, it fails to see how they were both right and necessary for the time in which they occurred (Herder Suphan 5:490; PW 282). It is, he says, as if an old man expected a baby, or a young boy, to live as he does, to think the same things, to value the same things, to take pleasure in the same things, as if we expected a baby or child or youth to be happy living a life suited to an old man (Herder Suphan 5:486, 489–90; PW 279, 282–3).

Herder also accepts the traditional idea that each people, and each stage of history, also goes through something like a life-cycle, involving a period of immaturity and growth, then a period of decline and decay, between them a time of flourishing, during which alone it fully enjoys the unique happiness proper to it (Herder Suphan 5:509; PW 296). In these cycles, peoples, nations, and whole ways of life and ways of thinking arise, mature, and then fade away to make room for others. This leads to the other main point—Herder’s “great theme”—that in the course of history a providential plan becomes visible, a necessary narrative structure of history, so that through all its changes “humanity ever remains only humanity” (Herder Suphan 5:511; PW 298).

According to Herder, the present age misjudges other ages on this score as well, by failing to appreciate how each earlier stage of history was necessary not only for its own time, but also for the entire process, and in that sense also necessary for the present age itself. Thus he accuses his contemporaries not only of condemning other ages by judging them according to the wrong standards, but also of failing to appreciate with gratitude how past ages were necessary to make possible their own way of thinking and its happiness. He insists, however, that it is not his aim to defend the way of thinking of other ages, but only to explain them (Herder Suphan 5:526; PW 309). Both praise and blame of the institutions of other times, he thinks, too often overlooks that what seems to us good and what seems to us bad are, in relation to the circumstances and way of thinking of that age, necessary to each other—as the slavery of the Helots was necessary for Spartan virtue (Herder Suphan 5:508; PW 295), and as periods of darkness were necessary to make our own enlightened age possible (Herder Suphan 5:525–6; PW 308–9).

The claims about *necessity* here are important to Herder, and he holds them in a very strong form. Later, in the *Ideas* he states as his “principal law”: “Everywhere on our earth whatever could be, has been, according to the situation and wants of the place, the circumstances and occasions of the times, and the native or generated character of the people” (Herder Suphan 14:83; *Ideas*, Book XII, Ch 6). This “law” would seem to be for Herder—to put it in Kantian terms—a regulative idea for the “explanation” he seeks. We would fully explain a people or an age when we see how its feelings and thoughts, its actions and values, its strivings and its happiness, are

all necessary to one another, and also necessary not only to the people's geographical and economic circumstances, but also to its place in the historical succession of ages through which human nature develops.

We can best understand Herder's critique of his own age if we view it as simply an application to the present age of the basic principles of his theory of history. The present age, like every age, is intolerant of others, and regards its way of thinking and its happiness as the only way of thinking, the only way to be happy. It is this thought, I suggest, rather than anything Herder found in the thoughts or writings of Enlightenment thinkers, that leads him to depict the Enlightenment as viewing itself with arrogant complacency as the final goal of history (Herder Suphan 5:559–60; PW 334–5). The other main point to notice is that in relation to his conception of the life-cycle of an age, Herder regards his own time as a period of decline and (as he repeatedly says) of “exhaustion” (Herder Suphan 5:538, 556, 582; PW 319, 333, 355). This is the deeper meaning of his metaphorical description of the present age as an old man. It colors his interpretation of it as a one-sidedly philosophical age, characterized by hyper-intellectualism, the separation of head and heart, a loss of faith in religion and of commitment to fatherland, and, in sum, what he calls its “mediocrity of soul” (Herder Suphan 5:583; PW 356) and its “human misery” (Herder Suphan 5:526, cf. 538, 550; PW 310, cf. 319, 328). For Herder, the high point of modernity was the practical flourishing of mechanical inventions in the seventeenth century, which formed the age's models of science, its politics, and its philosophy (Herder Suphan 5:536–7; PW 317–18). The high point politically came with the reign of Louis XIV (Herder Suphan 5:581; PW 354). The deepest mistake of the age is to think of its philosophy, its knowledge, its self-understanding, as the culmination of all history, since in Herder's view these are not a culmination of anything, but merely symptoms of decadence.

This also accounts for the fact that Herder ends his manifesto with an enthusiastic expression of hope for a new age, in which (as in all ages past), the achievements of the present age will be appropriated and at the same time transformed into something entirely new and higher. It is in this spirit that Herder praises Enlightenment ideals—knowledge, freedom, equality, sociability (Herder Suphan 5:575; PW 349). In their present form, to be sure, they often do as much harm as good, leading to shallowness, corruption, and misery. Their true meaning and value will come to be appreciated only in the future, when they will be seen as mere fragments of a larger whole, as what he calls “the fragment of life” that we presently are, after the plan of Providence eventually reveals them as tools of a larger purpose than we are now in a position to understand (Herder Suphan 5:584–6, cf. 5:513; PW 356–8, cf. 299). Herder's critique of his own age is simply a consequence of his acceptance of a form of historical optimism, of an objective natural teleology of history.

5.5 Herder and “Cultural Relativism”

Herder’s view of history and society, as I have already mentioned, has often been described as “cultural relativism.” This term is used in such varied, imprecise, and even confused senses that until we make it more precise, it would be just as pointless to deny that it describes Herder’s views as it is uninformative to apply it to them. But we should realize that such terms were not current in Herder’s day, and he never describes his own views using them. When we represent Herder’s views using terms like “relativism” and “historicism,” there is a serious danger that we will not only distort them but also miss parts of them that might help us decide what is right and wrong with the later views bearing such names.

“Relativism,” like many philosophical theses, but to a far greater extent than most, finds itself perched on a thin ledge between, on the one side, trivial truth (something that few, if any, have ever thought of denying), and on the other side, absurd falsity, self-contradictoriness, or sheer nonsense. Answering even the simplest questions often forces a relativist to choose between plunging into the one abyss or the other. Herder’s enthusiastic and undisciplined style of thinking and writing often opens him to charges of inconsistency, but I think there are some things we can note about his views that help exonerate him from the charges of incoherence or self-refutation that are often brought against various forms of relativism.

First, although Herder is aware how difficult it is to understand other times and other cultures, he never doubts that the kinds of claims he means to make about ancient societies, as well as about his own age, are *objectively true*. There is for him only one objectively correct understanding of any age—namely, the one suited uniquely to that age itself, the necessity of its relation to its geographical and historical conditions, and the necessary connection of its own way of thinking and acting. Any other understanding of it, such as one drawn from the standards and prejudices of a different age, is objectively false. Historical or cultural truth is not at all relative to the perspective of the observer, although our perspective may make that objective truth more difficult to know.

Second, Herder also rejects the “cultural determinism” that belongs to the theories of some more recent anthropologists, such as Ruth Benedict or Clifford Geertz. This is the view that some entity called a “culture” always decisively conditions all individual human thinking and provides the only criterion for the correctness of the thoughts and actions of individuals within the culture. On the contrary, for Herder the deepest truth lies in individuality, and the true measure of the happiness and virtue of individuals is always their individual life. Culture provides only the necessary context for that endless richness.

Herder's closest approach to what we now call "relativism" regarding ethical issues is probably his view that the character of human happiness, and consequently also of morality and virtue, varies fundamentally from culture to culture and age to age. But this view is best seen as a consequence of combining two other views:

Moral Eudaimonism (ME): Valid moral standards depend on the nature and means to the happiness of those to whom they apply.

Variability of Happiness (VH): There is no invariant human nature that determines what is best for all human beings. The content of human happiness varies greatly culture to culture, and even more from individual to individual.

ME is held by many philosophers, including Aristotle and the utilitarians—though eudaimonists certainly disagree among themselves about what happiness is and precisely how moral standards depend on it. VH is held by fewer philosophers, and seems to be rejected by Aristotle and other ancient eudaimonistic schools of ethics. (Aristotle recognizes the wide variety of opinions or *endoxa* about what happiness is, but is dismissive of most of them as merely views of "the many"; he thinks there is an objective truth about what happiness is that philosophy is capable of knowing (Aristotle 1095b–1096a).)

One philosopher in Herder's time who very explicitly embraces VH is Kant, who, however, clearly rejects ME. Nevertheless, it would be absurd to regard Aristotle or Bentham, or Kant, as sharing even partly in moral or cultural relativism. And it would seem strange that Herder should become a relativist merely by combining two already familiar views, neither of which is relativist. ME proposes a universal and purely objective standard for determining the content of morality. VH tells us that this standard applies very differently to different cultures, ages, and individuals. If relativism is simply the view that the same objective standard will apply differently to different people when the facts of their situations differ, then it begins to look trivially true, and something no one would ever think of denying. If relativism is the thesis that the combination of ME and VH entail that there is no objective truth about happiness or virtue, then the essence of relativism seems to consist in nothing but an elementary falsehood resulting from a patent *non sequitur*. It seems gratuitous uncharitableness to ascribe this elementary error to Herder, though many of his readers, as well as an entire tradition in cultural anthropology, may at times be guilty of it.

This is not an unfamiliar dilemma for moral or cultural relativism. William Graham Sumner, for example, most often states cultural relativism as the doctrine that the moral rightness of acts for members of a society is determined by the society's beliefs about what is morally right. At times, however, he claims that acts or

customs are morally right because they are well adapted to the needs and circumstances of people (Sumner was in many ways one of the crudest social Darwinists who ever lived). At those points he seems to treat the moral beliefs of a society as merely an empirically reliable guide to what has the property of being well adapted (Sumner, "Folkways," in Ladd, 1973, pp. 23–9). Similar shifts occur in the writings of other cultural relativists, such as Melville Herskovits (Herskovits, "Cultural Relativism and Cultural Values," in Ladd, 1973, pp. 58–78). That moral rightness consists in adaptation to a certain way of life seems quite close to Herder's view, but it is doubtful whether by itself it deserves the name "relativism." If it is suggested that such a non-relativistic view about moral rightness suddenly counts as cultural relativism whenever the theorist also believes there is radical variation between individuals or cultures in which acts have the proper relation of fittingness, suitability or adaptation, then it should be pointed out that it is highly misleading to think of moral or cultural relativism as a distinctive view about ethics or value, since when it comes to the fundamental ethical standard—fittingness, suitability, or adaptation to something—it agrees with a wide variety of ethical theories that no one would call relativist.

Herder's closest approach to relativism, however, seems to me to consist not so much in any view that he holds as in his awareness of a certain problem that arises when we become aware of the radical variation historical and cultural in ways of thinking. Herder perhaps states this problem most clearly in an interesting early fragment entitled (rather misleadingly) *Change of Taste* (1766). There, Herder observes that as soon as we become convinced on the basis of reasons that anything is true or good or beautiful, we straightway expect that everyone else will agree with us on the basis of those same reasons. But in fact we find (clearly this is one surprising result of Herder's approach to our understanding of alien ages and cultures) that others may regard the same things as false, bad, or ugly, also on the basis of reasons, so that (as Herder puts it) "truth, beauty and moral value is a phantom that appears to each person in another way, in another shape, a true Proteus who by means of a magic mirror ever changes and never shows himself as the same" (Herder Gaier 1:149; PW 247).

Herder realizes that this "contradiction," as he calls it, may lead us to doubt our own convictions, and even tempt us to become principled skeptics about everything. But he firmly rejects this reaction, both in *Change of Taste* and when the analogous problem arises in *This Too a Philosophy of History* (Herder Suphan 5:583; PW, 355, cf. Herder Gaier 1:151; PW 248–9). When in *Change of Taste*, Herder asks the question "Is not truth, fairness and moral goodness the same at all times?" he replies unhesitatingly in the affirmative, though he significantly admits that it is impossible to be entirely comfortable in answering this way (Herder Gaier 1:160;

PW 256). And in other places, such as a journal entry dated three years later, he flirts with the contrary answer (Herder Suphan 4:472).

Quite often, I submit, “relativism” is a name given to any pattern of thinking characterized by an awareness of the very real problem Herder is raising in *Change of Taste*, followed rapidly by a facile, complacent (and fundamentally dishonest) dismissal of the problem by offering oneself some set of skeptical sounding assertions that, when more closely examined, simply don’t make coherent sense. (If we want a precise definition of “cultural relativism,” that would be mine.)

Looked at from the standpoint of this problem, Herder can be called a cultural relativist to the extent that he acknowledges the problem, but I do not think he can be accused of its facile and incoherent dismissal, so he is only half-relativist (the honest and correct half, which leaves us in a state of intellectual dissatisfaction, not the dishonest and complacent half, which tries to run away from it).

Herder did, however, eventually formulate a solution to the problem, at least in regard to the variation involving different historical ages. It was not a facile solution, and it was also one with which, to his credit, he was never wholly satisfied. This is his theory of the necessary teleological development of humanity through historical stages. We reconcile the way of thinking that characterizes earlier ages with our own way of thinking by understanding those earlier ages as necessary not only to the place in history that we occupy, but also to the future development of humanity in accordance with a plan of Providence. In his historical manifesto of 1774, Herder’s position seems to be that we are capable of discerning that there is such a plan, but we realize that its end or goal is forever hidden from us, because the further development of the plan will take humanity beyond us into a future we cannot pretend to understand (Herder Suphan 5:513; PW 299). In the *Ideas*, Herder has found a name for the goal of this process—he calls it *Humanität*—a term he defines as “reason and equity in all conditions and all occupations of human beings” (Herder Suphan 14:230; *Ideas*, Book XV, Ch. 2). In other words, the resolution to the problem of historical variation is that the human species, through the secret plan of Providence, is striving ever forward towards precisely those ideals that were central to the Enlightenment.

One claim that is commonly ascribed to Herder is that different cultures or ages are “incommensurable.” (For example, see Beiser, 1987, pp. 143ff; Berlin, 1997, pp. 234ff). But I cannot find this term, or anything like it, anywhere in his vocabulary. Its use seems motivated mainly by a desire to massage Herder’s views in the direction of those the interpreter wishes he had expressed. Herder does, of course, think that each culture and time should be explained from within, as a unique development of organic forces, rather than as an instance of universal natural laws like those of Newtonian mechanics. He likewise rejects any standards of evaluation

that supposedly rest on a uniform and invariant human nature. So he thinks different cultures and ages are not “commensurable” in *those* ways. But the whole point of Herder’s philosophy of history is to enable us to construct a narrative progression leading from earlier times to our own, in which each culture and each age are understood in their own terms and yet also related to the progression of human history under the idea of *Humanität*. The latter relation is not one of incommensurability, but rather of a new kind of historical commensurability.

Herder’s counter-Enlightenment admirers are usually aware of this, but that doesn’t mean they have to like it. So when they come to the topic of *Humanität*, their stance is often to grumble, prevaricate, and condescend. “Perhaps,” admits Isaiah Berlin, “Herder did come to believe [in *Humanität* as a single uniting goal of history], or at least to believe he believed it” (Berlin, 1997, 234). When it comes to the central unifying idea of Herder’s mature philosophy of history, it seems that I differ from Berlin only in believing that Herder believed what Herder believed he believed. Or according to Michael Forster, Herder had doubts about *Humanität*, “just below the surface” (Forster, PW xxviff). Or maybe Herder had no such doubts after all, but we can still ignore *Humanität*—the unifying idea of his philosophy of history—simply because *we* have such doubts:

[Herder’s] philosophy of history is initially likely to seem striking and interesting mainly for its development of a teleological conception of history as the progressive realization of “reason” and “humanity”—a conception which anticipated and strongly influenced Hegel, among others. However, this conception is highly dubious on reflection, and is arguably *not* one of Herder’s main achievements in this area. (Forster, 2007).

My argument here is that Herder’s historical teleology is integral to his philosophy of history, hence also “arguably” inseparable from his main achievements in that area. (Isn’t it strange how the word “arguably” permits people to make directly contradictory claims, *arguably* with equal justification?) Along with Kant, Herder is the chief source for the great nineteenth-century theories of history in the German idealist tradition: those of Fichte, Schelling, Hegel, and Marx. This is an intellectual tradition for which no one needs to apologize.

5.6 Herder vs Kant

Herder and Kant did disagree about many things. Herder didn’t think there was such a thing as a priori knowledge, though (as it is with many empiricists, and especially pragmatists) this is more a matter of irresponsible emotive gesticulating than a well thought out epistemology. Herder, like many in the following generation of German philosophers, rejected Kant’s division of our theoretical faculties

into sense and understanding, and of our faculty of desire into reason and inclination. In matters of taste, he rejected Kant's classicism, and he favored an organic metaphysics composed eclectically out of doctrines from Spinoza and Leibniz that Kant regarded as enthusiasm.

In the philosophy of history, however, we have a fairly definite account of their differences in the form of the criticisms of Kant that Herder offered in the *Ideas*, and to which Kant replied in his review of the second volume of that work. These differences are real, but I do not think they are as fundamental as some would maintain they are. In fact, I think we will see that all the main disagreements concern the precise way in which we should understand the objective natural teleology of history—a teleology both thinkers equally accepted, though Herder perhaps gave it a more metaphysical, theological, and constitutive interpretation, while for Kant this teleology was always a regulative principle of reason in service of maximizing the comprehensibility of history to us. The principal *difference* between Herder and Kant, I would say, is that Kant, like many in his age, never fully comprehended or tried to practice the difficult method of understanding other ages that Herder advocated, outlined, and tried to execute. For the same reason, he never truly confronted the “change of taste” problem to which Herder's theory of Providence's plan of development toward *Humanität* offers, even in Herder's own view, I think, only a partial and less than fully satisfactory solution. But these most important differences are not disagreements so much as merely omissions on Kant's part to take up the most original and challenging ideas of his erstwhile student.

When it comes to actual disagreements in the philosophy of history, I suggest that the main points of conflict are three in number:

1. The role of the political state in the teleology of history.
2. The place of human happiness among the ends of nature (or Providence).
3. The understanding of earlier ages—especially the unhappiness found in them—as means to later stages in historical development.

In *Idea for a Universal History with a Cosmopolitan Aim* (1784), Kant argues that reason requires us to seek for an unconscious natural purposiveness in human history as a way of maximizing its theoretical intelligibility. The purposiveness he posits as a regulative principle is the indefinite development of the faculties of the species. This development, he argues, is seen empirically to occur through the mechanism of the unsociable sociability of human nature, which leads people to compete with one another, and by means of this competition progressively to enrich their faculties. The social antagonism that serves as nature's means, however, when it manifests itself in the form of violence among human beings eventually threatens nature's end. The end both of nature and of moral reason therefore requires the

establishment of a law-governed civil society capable of enforcing peace with justice among human beings. Nature's end is thus served by the tendency in history for human beings to perfect the form of a civil constitution administering justice, and this, in turn, Kant argues, will require establishing peaceful relations between states through a federation among them. In the course of explaining the difficulty of seeking a just constitution, Kant declares that the human being, in the social condition is "an animal who needs a master," yet the problem is that this master will also be an animal who needs a master, making the construction of a perfectly just constitution a task with no ideally perfect practicable solution (I 8:23).

In the second volume of the *Ideas*, Herder attacks Kant's proposition that "the human being is an animal who needs a master" as an "easy but evil principle." Kant replies that it is easy, because everywhere confirmed by sad experience, but he denies it is evil (RH 8:64). Kant does not mean, of course, that human beings are born for slavery, since the whole point is that they require a coercively enforced law to protect the external freedom that goes with their dignity as rational beings. The real disagreement concerns the role of the political state in the natural (or Providential) teleology of history. Kant regards coercion as an essential and permanent part of human life, owing to the unsociable sociability and hence the innate viciousness of human nature. Herder views the progress of humanity as resting on the victory of nobler and milder human predispositions, characterized by *Humanität*. We might see Herder as hoping for the eventual abolition or withering away of the state, whereas Kant entertains no such hope. This gap narrows, however, when, in the *Religion*, Kant comes to see the highest manifestation of historical teleology as taking place not in the coercive state but in the "moral commonwealth"—whose model is the religious community or church.

The second disagreement arises out of Kant's view that the natural teleology of history makes use not merely of social antagonism but also of human discontent as a spur to the development of our species faculties. For Kant, human happiness is an end of prudential and even of moral reason, but it is not an end of nature. Human beings were not put on the earth to be happy; their struggle to achieve happiness, though required by reason, is profoundly contrary to nature's plan for them. For Kant, the struggle to be happy is a struggle against nature (KU 5:429–31).

Herder agrees that perfect happiness or contentment will never be possible for human beings (Herder Suphan 14:333; *Ideas*, Book VIII, Ch. 5). But for him happiness is not a delusive will-of-the-wisp as it is for Kant, "Every living being [says Herder] rejoices in his existence; he does not inquire, he does not strictly examine, why he exists: his existence is to him an end, and his end is existence" (Herder Suphan 14:337; *Ideas*, Book VIII, Ch. 5). Herder regards Kant's notion that the plan of Providence would require human beings to be permanently *unhappy* as an impious indictment

of the Deity. For Kant, on the contrary, there are higher ends than happiness, and the moral worth of our person, rather than the happiness (the worth of our state or condition) is the correct measure of the meaning and value of our existence. Thus if human beings lived, as Kant imagines the inhabitants of Tahiti to do, merely for the pleasure of existing, without developing or exercising their properly human faculties, then Kant wonders “whether it would not have been just as good to have this island populated with happy sheep and cattle as with human beings” (RH 8:65). To prevent misunderstanding, we need to add here that of course Kant does not believe that it would be just, or even possible, for Europeans to “civilize” other peoples. With full Kantian severity, he condemns the unjust actions of European colonialism, in invading and exploiting the inhabitants of other parts of the world (EF 8:358–9; MS 6:266). But he does think that human beings everywhere give meaning to their existence only by developing, according to their own lights, the capacities of the human species, thus fulfilling the purpose of both nature and reason.

This issue is closely related to the third one. Herder finds absurd Kant’s idea that the end of nature for the human species, namely, its indefinite development of its rational predispositions, should be thought of as fulfilled only in the open-ended future of the human species as a whole, never in the happiness of individual human beings at any given time.

If someone said that not the individual human being but humankind is to be educated, then he speaks unintelligibly for me, since kind and species are only general concepts, except only insofar as they exist in individual beings.—It is as if I spoke of animality, minerality and metality in general and adorned them with the most splendid attributes, which, however, contradict one another in single individuals! (Herder Suphan, 13:345–6; cf. RH 8:65).

Kant’s reply is that when he speaks of the end of nature being fulfilled only in the human species as a whole, he does not mean the abstract general concept of the species, but rather something particular and concrete—“the *whole* of a series generations going (indeterminably) into the infinite” (RH 8:65). In other words, Kant thinks the end of nature in history, the endless development of our species capacities, is fulfilled in the same way that Herder thinks the end of nature is fulfilled in the human species through the endless progress toward *Humanität*. The disagreement is over whether it is a fitting way to think of the plan of Providence to regard the happiness (or rather, for Kant, mainly the discontent) of all individuals, and of every generation, as nature’s means for fulfilling an end that no individual and no generation enjoys. You might say that Herder is afraid that on Kant’s view of history, God is guilty of violating the Kantian principle that humanity in everyone’s person should be treated as an end, not merely as a means. “Not a thing in the whole of God’s realm,” Herder declares, “is *only* means—everything is *means*

and *end* simultaneously” (Herder Suphan 5:527; PW 310). Of course for Herder, the only way individual human beings can be an end is by being happy, rejoicing in their own existence, whereas for Kant, happiness is only a conditioned good, and the worth of human existence is measured by human perfection, especially the moral perfection of will. Yet Herder himself has a problem of the same kind, owing to his doctrine that there is a life cycle for every age and every people, with its true happiness attained only during the period of its maturity (Herder Suphan 5:511; PW 298). It is not clear how he can clear himself of the charge that those who endure less happiness during periods of growth and decline are not being treated by Providence as mere means to the end of the happiness of those who enjoy the full flourishing of their age.

The whole issue, however, seems to matter only if you are preoccupied with the problem of theodicy. Once we begin to think of the human species in history as struggling to create meaning in the context of a nature that sometimes treats us kindly, more often cruelly (its final act toward each of us will be murder), but a nature that is always fundamentally blind and indifferent to human concerns, we no longer expect nature to obey the Categorical Imperative, and the point of controversy here between Herder and Kant just disappears for us.²

These real points of dispute between Kant and Herder are interesting and significant, but not nearly as important as the points of agreement that made them all possible. If our aim is to show Herder as the superior thinker, then we would be wise not to portray him as a critic of Kant or of the Enlightenment, but rather to focus on his positive accomplishments, which (however he may have understood them) in fact enabled the Enlightenment’s values and vision to be extended to a richer appreciation of the full wealth of human life and history.

In his later writings (starting about 1789), Kant offers three “maxims” for our use of reason (or understanding):

1. Think for yourself.
2. Think from the standpoint of everyone else.
3. Think consistently.

² For most of us, anyway. A few seem to think that just to *look* at history as progressive is to commit wickedness, by making oneself complicit in history’s evil deed of treating past generations as mere means to future ones. One who thinks this way is Hannah Arendt: “In Kant himself there is this contradiction: Infinite progress is the law of the human species; at the same time, man’s dignity demands that he be seen (every single one of us) in his particularity, and, as such, be seen—but without any comparison and independent of time—as reflecting mankind in general. In other words, the very idea of progress—if it is more than a change in circumstances and an improvement of the world—contradicts Kant’s notion of man’s dignity. It is against human dignity to believe in progress” (Arendt, 1982, p. 77). Similar charges against Kant are found in Galston, 1975, pp. 231–5; Stern, 1986, pp. 535–9, and Dupré, 1998).

To me, these maxims exemplify the spirit of Kantian philosophy, and of the Enlightenment, as profoundly as any thoughts he ever expressed. I view Herder's great contribution as, in effect, a remarkably original expansion on the second rule. Kant always regarded all three rules as extremely difficult to satisfy, and impossible for flawed and limited creatures like ourselves ever to follow perfectly. Herder showed us a new dimension of difficulty in following the second rule, of which Kant and most of his contemporaries had little or no conception, but he also opened up for us a new way of thinking about culture and history that makes possible a new kind of striving to follow it.

5.7 Faith in Historical Progress as a Rational Faith

Herder and Kant both seek to comprehend history by supposing that historical events promote a certain end (or ends). The ends they use are different in the two cases, so their theories of history differ materially in certain respects. But I think the grounds, both theoretical and empirical, that they use to justify their theories are quite similar. For Herder, the end of history is *Humanität*—"reason and equity in all conditions and all occupations of human beings." This goal emphasizes both the endless variety of human institutions, experiences of life, and ways of self-understanding, and also a tendency of different human beings to understand this variety, rationally and equitably, as a valued expression of different sides of human nature. For Kant, the basic tendency in history is the endless development of our species predispositions through unsociable sociability, which can continue (after the coming of civilization) only through progress towards the idea of a law-governed civil society administering justice and then (at a certain stage in the development of states) through progress towards an international organization of states maintaining a just peace between them. Neither philosopher maintains that progress toward these ends is smooth or uninterrupted. Kant even holds that the idea of a law-governed civil society is to be employed by historians "to show how far humanity has approached this final end in different ages, or how far removed it has been from it, and what is still to be done for its attainment" (Ak 8:468). In other words, there is no dogmatic optimism about the course of human events.

In both Kant and Herder, a teleological view of history can be seen as a response to two simultaneous demands of rational inquiry: one theoretical, the other practical. On the one hand, as rational inquirers into human history, we seek some kind of systematic comprehension of the facts, by showing how they exhibit certain underlying tendencies. On the other, as historical agents, we seek to comprehend our own actions as part of an ongoing historical pattern in which we may

contribute towards some worthwhile end, which history exhibited before our arrival and can be expected also to exhibit after we are gone. Both motives can be understood as rationally required if we are to make sense of history, whether as rational inquirers or as purposive agents. As rational inquirers, we seek for something to unify the events of history into a meaningful pattern. Kant argues, in the *Critique of the Power of Judgment*, that in looking for a systematic order among natural phenomena that cannot be explained by mechanistic causal laws, it is rationally required that we employ the concept of a natural end. This is what Kant does with his theory of the development of our species predispositions through unsociable sociability, and what Herder does when he seeks a connectedness among different peoples and ages, ultimately employing the idea of *Humanität*. As agents, we seek to understand our own actions as contributing to a meaningful historical process. We seek for historical trends or tendencies that harmonize with our human strivings for justice, peace, equity, and reason, so that we may both understand our agency better and use it to contribute to something in history we consider worthwhile. These are not merely optional aims for us insofar as we are rational knowers and purposive agents who also recognize ourselves as historical beings. If we choose not to look for meaningful purposive patterns in history, and not to strive for a better future for our species in history, this is in effect to give up on our own rationality as historical beings.

Of course, for both purposes we must be flexible—as both Kant and Herder are—in our conception of the ends we ascribe to history. Whatever ends we see in history, and whatever historical ends we set for ourselves, we must conceive them only provisionally, remaining open to the empirical data and to our own fallibility in the way we specify them. *Humanität* is a conception that might be specified in endless ways, and it is our task both as knowers and agents to adapt our conception of it to fit our knowledge both of what has happened (and is happening) in history and of what ends are worthwhile for human beings to pursue in history. Likewise, the content of human predispositions is not something fixed in advance, just as the idea of a wholly just civil society is something constantly evolving as we learn more about human social life and about the requirements of right itself. But without some conception of this sort, as historical inquirers we would have to abandon the task of making systematic sense of history, and as agents we would have to cease to think of ourselves as part of a historical progression at all. Moreover, if we undertake both the theoretical and the practical task of making rational sense of history, we are bound to try to harmonize the two projects, so that the sense history makes to us as agents fits into the meaning we comprehend in history as knowers.

We may, of course, choose to despair of the theoretical project of comprehending history, as well as the practical project of contributing to a collective project

of the human species through time. It would no doubt be possible simply to give up on both projects, based on the kinds of horrors and disappointments that have disillusioned many historical optimists during the twentieth century. This would seem to be precisely what has been done by those who distance themselves from the entire Enlightenment conception of historical progress. The crucial point to make, however, is that no set of empirical facts could actually justify such an act of despair. It can be made sense of only as a certain irrational emotional decision, prompted perhaps by the failure of certain specific (and always revisable) conceptions about where history is and ought to be heading. Instead of revising these conceptions when the facts indicated the necessity of so doing, some people found this task too hard or too emotionally costly, perhaps owing to their irrational attachment to these provisional conceptions. So rather than give them up, they chose to abandon the entire rational enterprise of understanding history and orienting one's actions towards it. In contrast to this, Kant, Herder, and the entire Enlightenment tradition of thinking about history insists that we can rationally abandon neither the theoretical enterprise of comprehending human history nor the practical enterprise of contributing to a better future for humanity. This tradition insists instead on a sober perseverance in these enterprises, and an unwillingness to give up either.

This kind of situation, and the dilemmas it poses, is illuminated by thinking about it in terms of Robert Adams' conception of moral faith. For Adams, "faith" is the name for a response to a situation in which we believe something that is closely bound up with a project of some kind to which we are committed, but where there are also grounds for a reasonable person to doubt these same beliefs. "Faith" is the response that remains true to the project by holding to the beliefs it presupposes, in spite of the temptations to doubt these beliefs and give up on them and on the project (Adams, 1999, pp. 373–89). That faith occurs only where some degree of doubt is reasonable is the reason that Paul Tillich, for instance, claims (perhaps with an air of paradox) that doubt is a necessary element of faith. It is also the point being made by certain religious critics of dogmatic fundamentalism when they say: "The opposite of faith is not doubt but certainty." The Enlightenment conception of history, as exemplified in Kant and Herder, is an attempt to make sense of human history in the face of obvious reasons to doubt that it makes sense. Thus their philosophies of history exhibit *faith* in precisely Adams' sense.

Adams points out that faith, properly speaking, is not directed merely to propositions, or forms of words, but "a stance in relation to something larger" (Adams, 1999, p. 374). In this discussion, Adams is discussing faith in *morality* and in moral ends. Religious people have faith in God, or in God's goodness, his promises, his love or his mercy. Enlightenment historical faith is faith in human

reason, as capable of comprehending our history as something meaningful, and in ourselves as rational beings, as capable of building a better human future. This Enlightenment faith can be a religious faith—both Kant and Herder relate their philosophy of history to the idea of divine providence. But it can also be a purely secular faith, and faith itself (as Adams describes it) need not have anything religious about it at all.

Adams notes that faith in something—in morality, or in divine purposes—is entirely compatible with being open to the evidence to revising your beliefs about what these are (Adams, 1999, pp. 383–4). We have just noted that Herder and Kant both conceive the ends of history in such a way that they invite such empirically sensitive and revisionary attitudes, so it is a caricature of Enlightenment views of history to see them as blind to the facts or “aprioristic” in some bad sense of the term (implying that they ignore empirical evidence). More generally, Adams rightly emphasizes that faith need not be an unreasoned belief, still less an irrational one (Adams, 1999, p. 375). In fact, he describes it in Aristotelian terms as the virtue that lies between the vices of credulity and incredulity (Adams, 1999, p. 374). I think this is true of faith when it is a virtue, though I think faith is not always a virtue—for we can misplace our faith, by persisting in projects we should abandon and holding beliefs we should give up. In fact, I think faith remains a virtue only when it is fully rational. I would insist that faith, whenever it is a justified attitude, must be fully consistent with a moral principle that I agree with, which is often called “evidentialism”: the principle that our beliefs must always be proportioned to reasons or evidence. Some people use the word “faith” to refer only to beliefs that violate the evidentialist principle—as in Mark Twain’s wry definition: “faith is when you believe something you know damn well ain’t true.” But sometimes we have reason to doubt our beliefs but even better reason to stand by them, despite our temptation to give up on them. To do this is precisely to exhibit *faith* in Adams’ sense. Faith, whenever it is rational and a virtue, is therefore even one of the demands the evidentialist principle makes on us.

Thus insofar as faith is the virtue lying between incredulity and credulity, it involves beliefs we have some reason to doubt, but also reason to stand by, as part of our commitment to a project (theoretical, practical, moral) of some kind. Adams points out that not all the beliefs we hold are to be thought of as hypotheses to be tested and perhaps falsified (Adams, 1999, pp. 382–3). This is especially true in the case of what Kant calls “regulative principles of reason,” or beliefs involved in the pursuit of ends we are rationally required to pursue. The project of making rational sense of the world, and the regulative principles involved in it, are not subject to empirical falsification. Neither is the project of giving meaning through our actions to our own life or the collective life of humanity, or, therefore, the belief

that this is possible for us. I suspect that Adams may think that there is some refutation (or at least qualification) of evidentialism hidden in this last point. But I do not. Evidentialism, as I interpret it, tells us to be responsive to reasons and evidence in forming, maintaining, and revising our beliefs. (I defend a version of evidentialism in Wood, 2002, Chapters 1–2). The regulative status of some of these beliefs, or their involvement in practical projects we should not abandon, seems to me merely one species of these reasons.

Faith, when it is a virtue, never has anything to do with wishful thinking—with the sad, or comical, or dangerous, human tendency to believe something because we want it to be true or because believing it is true makes us feel good. Wishful thinking is always an irrational way of forming beliefs, showing disrespect for ourselves as rational beings, and the beliefs formed in this way are usually erroneous and often dangerous. Of course, if we set ourselves some end, and hold the belief that our end is possible of attainment, then we probably also want it to be true that our end is possible. But here the direction of our motivation is just the opposite of that involved in wishful thinking. Wishful thinking believes something because it pleases us to believe it. In rational faith directed to a rational end, we believe our end is possible because we have rationally chosen to pursue it. Having this belief might also please us, but this pleasure is only an incidental result or by-product of the rational belief, not its cause. (If it is really the cause, then we are talking about something other than, and less respectable than, rational faith.)

Wishful thinking, however, is far from being the only pattern of irrational belief formation. People are also subject to what we can call “fearful thinking.” Their dread of an outcome sometimes makes them find it more likely, and this sometimes leads them into irrational patterns of behavior as they become obsessed with dangers that have been exaggerated by their imagination. Then they flounder about among emotionally charged ways of coping with the dangers. The dangers may be real, but in their fear they have lost their rational grip on how to deal with them. Faith is precisely the proper rational response to fearful thinking. This makes faith, properly placed, entirely rational, and even an indispensable resource of human rationality.

Fearful thinking, for example, had a fateful role to play in the American reaction to the events of September 11, 2001. It made people think that Islamic terrorism is a greater threat to them than it really is. It led some people to think that an unprovoked military conquest of Iraq might be the only way of preventing weapons of mass destruction from falling into the hands of terrorists and being used against U.S. cities. It caused Americans to abandon their commitment to principles of human rights, to engage in the indefinite detention of certain individuals without charge or trial, and in acts of torture. Fearful thinking caused Americans to lose

faith in themselves, in principles of international law and due process of law. The same course of events also shows that when people lose faith in what they should have faith in, they often replace this lost faith with a misplaced faith in something they should not have faith in. For when the fearful events of September 11 caused Americans to lose faith in what is decent in themselves, it led them to put their faith in the wrong things—in acts of military aggression, vengeance, and injustice, and in a corrupt and incompetent political regime, from whose misdeeds we are still suffering, and from which the world is probably going to suffer for many years to come.

The Enlightenment conception of history, as exemplified in different ways by Kant and by Herder, is in this sense the result of an act of rational faith. Those who have rejected it on the basis of twentieth-century events such as the First World War, the Holocaust, or the failure and collapse of Eastern European communism, exhibit the kind of historical despair to which Enlightenment faith is exactly the needed response. Religious objections to Enlightenment optimism, for example, are often in large part expressions of despair over modernity, or humanity, or reason. To the extent that their God is the *totaliter aliter* of all these, it has to be reckoned a superstitious idol, or even an evil god. Some postmodernists are disillusioned Marxists, who would have done better to have kept faith with Marxism, while perhaps revising some outdated aims and brittle doctrines to fit reality, instead of letting historical despair be their last act of stubborn allegiance to an irrational dogmatism. Others are simply followers of an aesthetic fashion of despair, nihilism, and ironic detachment from all serious purposes in human life. In contrast, Enlightenment faith refuses to let the difficulty of making sense of history, or the painful frustrations caused by our disappointed historical hopes, push us into the despair of giving up on human history as a subject matter for rational comprehension or on ourselves as historical agents.

People who despair of the Enlightenment view of history need not despair of everything, of course, as many of them will be eager to point out. But they are despairing of rational comprehension of a human endeavor considered as a whole, and of thinking of themselves as rational participants in it. Often people who do this try to portray it as merely a form of epistemic modesty, as though philosophical reflection on history were all by itself a kind of metaphysical extravagance. And they might try to portray their denial of it as mere modesty and sobriety, rather than as a despair of anything. I hope they will forgive me for not being convinced by such flattering self-portrayals. For once an intellectual project, such as the Enlightenment attempt to comprehend history (and its various nineteenth-century descendants—which all remain children of the Enlightenment, however much they may self-deceptively try to think of themselves as superior to it or as “going

beyond" it), there is simply no way of turning your back on it without in effect despairing of it, and of the functions of both theoretical and practical reason that it represents.

Kant's sober insistence that we must look for rational purposiveness in history, and for a set of moral tasks arising out of our condition as historical beings, is a form of rational faith. Herder, too, as I have argued, did not repudiate the Enlightenment in this respect, but in the end kept faith with it. His greatest contributions were instead towards deepening and enriching it—in ways of which, like the radical aims of the Enlightenment themselves, we still cannot see the end.

6

Leaving Consequentialism Behind

6.1 “Consequentialism” vs “Deontology”

It is a commonplace in moral philosophy that there are two basic sorts of moral theories: consequentialist and deontological. Consequentialist theories are supposed to base actions, reasons for action, and even theories of such reasons, solely on the value (good or bad) of the results (expected or intended), whereas deontological theories are supposed to tell us what actions to perform, and why to perform them, entirely on the basis of something other than the goodness or badness of their intended or expected results—on some accounts, supposedly on the basis of no value whatsoever. The difference is often presented in terms of people’s differing reactions to (or, as it is said, their “intuitions” about) certain examples. For instance, a consequentialist is supposed to approve of an action whenever it results in saving five lives, even though it would involve killing an innocent person, whereas a deontologist is supposed to condemn any action that results in the death of an innocent person, even to save a great many lives, because such an action is always wrong, no matter what the consequences.

Like many commonplaces (both in philosophy and outside it), this one won’t hold water. Would someone who considers the good or bad consequences of actions have to reach different conclusions from someone who considers other kinds of reasons? I doubt it, for reasons I will return to at the end of this chapter. Then too, neither theory, as commonly understood, makes much sense. To reject the good or bad consequences of an action as reasons for or against it would be as silly as ignoring on principle all the other reasons there might be. Both theories, as they are often invoked, seem to be caricatures, almost as stupid as the fantastic examples about which our “intuitions” are supposed to offer reasons for accepting or rejecting them. In relation to these “intuitions,” consequentialists look foolishly

shortsighted about the consequences of actions and blinkered in their ways of evaluating them, while deontologists appear pigheaded and inflexible about moral rules. Each theory, in other words, as it is often depicted, seems to be conspicuously stupid about the very thing that is supposed to matter most to it. When used in this way, both theories seem to be straw figures, designed to make some other alternative (such as no theory at all) look good by comparison. It seems to me a point in favor of ethical theories (both consequentialist and deontological) that the opponents of ethical theory have to resort to such desperate caricatures in order to make their nihilistic position seem plausible.

My main thesis, however, might seem to be in tension with what I have just said. It is that when it comes right down to it, deontologists are more right than consequentialists. That thesis (plus an allusion to the provocative title of a well-known paper by Barbara Herman) is how this chapter got its title (Herman, 1992, pp. 208–42). But provocative titles must never be taken too literally. I won't be trying to get you to look forward to a golden time when we can afford to ignore consequences in deciding what to do. Nor am I saying that when people reason consequentialistically about what to do, they are necessarily making any mistake. What I mean is that reasons for action other than those involving consequences go deeper, and even help us to explain why we should care about consequences when we should. My claim is that consequentialist reasoning needs a rational (at least implicit) grounding, and that grounding is deontological in nature. To put it briefly: we ought to care about the consequences of our actions only because we care more fundamentally about acting as we should.

The initial appearance, I admit, may be just the opposite: Bentham can be forgiven for thinking that those who have regard to good or bad consequences are thinking rationally about what to do, while people who think about reasons for action in ways not grounded in the value of consequences are merely expressing irrational sympathies and antipathies (Bentham, 1948, pp. 13–16). Bentham realized that there can be rationally justified rules (or laws) against actions of certain kinds, but he thought such rules needed to be justified by the consequences of following or promulgating them, because the ultimate reason for an action cannot lie in the action itself but in what the action is expected to bring about. I admit it might seem on the surface that things are the way they seemed to Bentham. But I will argue that the deeper truth turns out to be just the opposite.

6.2 What is Consequentialism?

It is sometimes suggested that consequentialism is a view solely about the rightness of actions—that (roughly speaking) an action is right if it produces the most

good, i.e. the best consequences. (A good statement of consequentialism in this sense is provided by Philip Pettit in Baron, Pettit, and Slote, 1997, pp. 92–174). It is sometimes inferred from this that consequentialism involves no determinate views about what the good is—that a consequentialist can simply leave it open how “the best consequences” are to be determined. But I think both these claims look at matters the wrong way. As Marcia Baron has pointed out, ethics deals with many questions besides that of the rightness of actions, and consequentialism involves a distinctive position on the nature of social policies, virtues or (as with Mill) rules of action, hence not only about the rightness of actions (Baron in: Baron, Pettit and Slote, 1997, pp. 239–244). Consequentialism, as I will understand it, is the position that all practical reasons relevant to morality either consist in, or can be reduced to, reasons provided by the value (good or bad) of states of affairs, considered as the possible outcomes of actions, where actions are considered as means, or causal contributors, to attaining or avoiding these states of affairs. That position certainly *is* committed to some determinate views about the nature of the good. It takes the good (and the bad) to consist in values instantiated by *states of affairs* as the *possible outcomes of actions*. (I suspect some do not recognize this substantive commitment because they lack the philosophical imagination needed to realize that there could be any alternative to it.) Consequentialism may not be strictly committed to, but it does also tend towards, certain views (to be discussed later) about the formal structure of the kinds of goodness and badness that such states of affairs might possess.

Here is a statement of the idea behind consequentialism that is as clear and cogent as I know of. In the brief opening chapter of *Utilitarianism*, Mill says: “All action is for the sake of some end, and rules of action, it seems natural to suppose, must take their whole character and color from the end to which they are subservient” (Mill, 2001, p. 2). Actions, whatever else they may be, are conceived as means toward ends beyond themselves—states of affairs to be brought about through them. This offers us a set of reasons for acting that belong essentially to the nature of action and specifically concern the consequences that we expect and intend when we act.

6.3 What is Deontology?

For the basic idea behind deontology, I offer the opening sentence of Part I of Fichte’s *System of Ethics*: “It is asserted that in the mind of the human being there occurs an absolute constraint to do certain external things entirely independently of external ends, merely and solely so that they happen; and to refrain from others, just as independently of ends external to them, merely and solely so that they

remain undone" (Fichte, SL 4:13). The crucial idea behind deontology is that actions are always to be valued (and disvalued) fundamentally for themselves, and never only as ways of reaching (or preventing) possible consequences, regarded as their effects.

Mill's and Fichte's statements both represent important insights about the structure of action, which are decisive in relation to ethical theory. But it is important not to confuse these insights with claims about which thoughts it would be most appropriate for people to have in the moment of action. I pause for a bit to dwell on this important point, because it is the source of many bad arguments against particular moral theories, and against moral theory in general.

The thoughts that are appropriate at the time of action might vary greatly with circumstances, depending on the reasons that are most salient in the heat of the moment, or on the thoughts that a virtuous person would have when confronted with precisely this situation. The suitable thoughts at that moment, however, might not be a good clue to the fundamental rationale for what they are doing, or what they should do. When you are concerned to help a friend who is in trouble, you would be oddly self-centered if you were dominated by the deontologist's thought: "This is an action worth doing for its own sake." To be occupied with that thought in the moment of action would make it seem as if you didn't really care about your friend at all, but only about your own virtuous agency. But it would be equally odd, and no less unattractive, if you were to have the consequentialist thought: "In general, things go better for people if friends help each other out, so in order to contribute to the greatest overall good, I should help this person now." If you could not be motivated to help your friend without deriving your concern from some general consequentialist principle, we might wonder whether you are capable of any genuine friendship at all. (This unjust charge against consequentialism is found, for instance, in Wilcox, 1987.)

Probably the most famous argument of this kind, mounted against moral theory in general, is Bernard Williams' example of the man who, in rescuing his wife from drowning, is said to have "one thought too many": "The idea that a moral principle can legitimate [the preference to save one's wife] ... provides the agent with one thought too many: it might have been hoped by some (for instance, by his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations like this it is permissible to save one's wife" (Williams, 1981, p. 18). But used against any particular moral theory, or against moral theory in general, these are bad arguments. The point is that while such theoretical thoughts might be superfluous (or worse) for someone to have in the moment of action, they may not be the least bit superfluous or out of place when we are engaged in the more reflective business of theoretically systematizing,

explaining, spelling out, and justifying the ways people act and the reasons they have for acting in certain ways. Thus a person can be a genuine friend even if, while doing ethical theory, they have the deontological thought that the act of helping is worth doing for its own sake, or the consequentialist thought that things go better when friends help each other out. And the same is true for the man who might, in a cooler hour, set about providing a theoretical account (whether consequentialist or deontological) of the reason why it was right for him to save his wife.

6.4 Both Insights Together—in Aristotle

Both the consequentialist insight and the deontological insight are to be found side by side in Aristotle's *Nicomachean Ethics*. The insight expressed by Mill occurs in the very first sentence: "Every craft and every line of inquiry, and likewise every action and decision, seems to seek some good; that is why some people were right to describe the good as what everything seeks" (Aristotle, 1999, 1096a1–3). (I ignore for the moment the apparent fallacy of composition here, though presently I will come back to it.) Fichte's insight is found even more repeatedly in the *Nicomachean Ethics*, where Aristotle says that virtuous actions are always done for their own sake (Aristotle, 1999, 1095a20, 1105a32). In fact, Aristotle even seems to extend this thesis beyond virtuous action to action (*praxis*) generally, to every decision or choice (*prohairesis*) (Aristotle, 1999, 1144a19, 1174a7–8), and even to every activity (*energeia*)—even to seeing, remembering, and knowing (Aristotle, 1999, 1174a5–8), at least when these are considered as human activities, rather than the mere sense perception of beasts (Aristotle, 1999, 1139a20). Human beings, unlike non-rational creatures, are aware of themselves as acting when they act, and they value their action for its own sake. Aristotle distinguishes action (*praxis*) from production (*poiesis*). Action, by its very concept, is valued for its own sake, while production is not; *praxis* is actuality (*energeia*) at every moment complete in itself, while production is mere motion (*kinesis*), never complete but always tending to something beyond itself (Aristotle, 1999, 1139a35b4, 1140b6, 1174a19–b5). This is Aristotle's reason for thinking that when it comes to the good life, money-making is a non-starter; money is merely useful for some other end; making it could not be choiceworthy for its own sake (Aristotle, 1999, 1096a7).

If we don't immediately recognize Fichte's expression of this idea as the same as Aristotle's, that is probably because Fichte speaks of the mind's being *self-constrained* to do what it does for its own sake. Fichte thinks that what we most value doing rationally and for its own sake, what we most deeply and most *freely* want to do, is not something that we do easily, spontaneously, or naturally, but something we have to constrain ourselves to do, overcoming inveterate corrupt

propensities that tempt us away from it. This deeply Kantian thought—acknowledging the corrupting influence of modern society on the human will—may make it harder to recognize Aristotle’s insight in modern dress; but it is nevertheless the very same insight—expressed in a way that is better suited to us who are condemned to live in modern capitalist society. We are less naïve about the meaning and effects of social inequality, hence better acquainted with, and more sensitive to, our own self-inflicted misery and depravity.

6.5 Deontology and the Ends of Action

Of course, one reason to value an action, even to value it for itself, might be that it has certain consequences, namely, those that accord with one’s ends. For it is not only part of the concept of an action that every action is done for its own sake—or as Fichte puts it, independently of ends external to it—but also (the point Mill insists on) that every action has an end beyond itself, a state of affairs to be brought about by the action. But if it is part of the concept of an action that it has such an end, then is the end internal to the action or external to it? The uncertainty here is due at least in part to the fact that the concept of an end is ambiguous: it can be taken in a wider or a narrower sense. The wider sense is indicated by Fichte when he says that “I posit myself as free insofar as I explain a sensible acting or being as arising from my concept, which is then called ‘the concept of an end’” (SL 4:9). An *end*, in this wide sense of the term, is no different from a ground or reason for acting, whatever it might be. This is well expressed by the German word for end: *Zweck*. The common expression “*Das hat keinen Zweck*” should not be translated into English too literally. “That has no end” makes you think of the series of natural numbers, or the calamities we Americans bring upon ourselves when we elect Republicans to political office. What the sentence really means, however, is: that’s pointless or futile; there is no reason for it. The *Zweck* of an action may, but it need not, consist in a state of affairs to be brought about by the action.

After noting that an action is valued (or disvalued) for its own sake, we may probe either an agent’s or a theorist’s reasons for holding that it is; and these reasons are heterogeneous. And it is an essential advantage of deontology over consequentialism that the reasons for valuing actions for their own sake are more encompassing than the reasons for valuing states of affairs (as possible consequences of actions). If we look at Aristotle’s reasons, for instance, for considering some ethical action “fine” (*kalon*), these turn out to be different in different cases. Sometimes consequences count, sometimes motivations, sometimes conformity to or violation of principles, sometimes it is the way an action manifests an admirable or shameful trait of character. Sometimes an action is valued or disvalued

for its own sake based on more than one of these kinds of reasons. It would be an error to think that just because we can cite a reason why an action is valued for its own sake, we can ignore the fact that it is so valued, or can simply replace it in our thinking by the reason or reasons why it is so valued (such as a consequence-based reason). For not every reason why we might do or refrain from an action counts as something that makes the action valued or disvalued for its own sake. The main claim of deontology, as I am understanding it here, is that it is basic to rational agency to focus attention on the value an action has for its own sake, rather than on the reasons we may have for valuing it this way.

Fichte thinks that every action, whatever its *Zweck* might be of a non-consequentialist nature, *also* has a *Zweck* in the narrower sense of a possible future state of affairs to be brought about by it. Every action should be done for its own sake, but every action *also* aims at a result to be achieved in the world. “[The fact that] I find myself acting efficaciously is possible only under the condition that I presuppose a concept projected [*entworfenen*] by myself, which is supposed to guide my efficacious acting and in which the latter is both formally grounded and materially determined” (SL 4:9). An end, in other words, is a concept involved in the agent’s acting that “formally” provides the ground or reason for action; this concept “materially” determines the subject, projecting the I onto the object or state of affairs to be brought about. Both the action as end in itself and the concept projected by the action are internal to the action itself. Thus when Fichte speaks of “absolute constraint to do certain external things entirely independently of external ends, merely and solely so that they happen,” he does not mean to exclude (under the heading of “external ends”) those contained in the self-projection which is that very action itself. He means to exclude only ends external to the concept of the action—such as the satisfaction of desires falling outside the rational desire constituting that self-projection which is the action itself.¹

Fichte’s notion of “projection” was, of course, later taken up by the existentialists, Heidegger and Sartre, when they held that the very being of *Dasein* or *réalité humaine* is the projection on one’s own future. This idea too captures the deontological insight found in Aristotle. It reveals that the directedness of an action to a goal or material end (an end in the narrower sense) is thought of as the *being* of the agent, a *determinate way for Dasein or human reality to be in the world*. In

¹ Kant draws a related distinction when he separates an “end to be effected” (*zu bewirkenden Zweck*) from an independent end (*selbständigen Zweck*), such as an end in itself (G 4:437). Kant too thinks that every action, by the very concept of action, has an end to be effected, in relation to which the action is thought of as a *means* (G 4:427). But an end in the widest sense is anything “for the sake of which” we act, and in acting we may also act for the sake of a person (oneself or another), considered as an end in itself, just as we may also act for the sake of the moral law, when we act from duty (G 4:390).

projecting itself on such a goal, and valuing its realization, the agent also constitutes its own being as an agent, which it thereby values for its own sake.

Fichte's notion of action as "projection" relates to a number of issues in his philosophy—to his conception of the I's absolute freedom, for instance, as well as to the temporal structure of action. But here I am interested in it only insofar as it makes the point that the deontological insight that actions are to be done or omitted for their own sake does not preclude (but on the contrary, explicitly embraces) the idea that actions aim at objects, alterations, or states of affairs in the world, that this is essential to the structure of action, and it is this purposiveness of action that may constitute precisely what makes an action good or bad for its own sake.

6.6 Consequentialism and Human Flourishing

If action is part of human flourishing, and every action is to be valued for its own sake, then it follows that there is something problematic about the following description (which was offered to me in discussion by Richard Boyd): "the consequentialist position maintaining that people should help one another to flourish." This describes a *consequentialist* position, however, only if it is understood in a rather special (and I think quite problematic) way: namely, it must regard human flourishing as consisting of nothing but the existence or obtaining of certain states of affairs as the possible outcomes of human actions, and it must value as human flourishing only the desirable properties of those states of affairs. If, however, Aristotle is right that human flourishing consists in activities valued for their own sake, then Boyd's formula is not consequentialist at all. Moreover, if the actions of helping one another to flourish are thought of as having a value for their own sake, and not merely for the sake of certain states of affairs as their outcome, then this formula is not even compatible with consequentialism, even if some of what is thought and done in the course of helping people to flourish might consist in engaging in consequentialist thinking and performing actions with a consequentialist rationale.

What I believe misleads us here is a certain superficial picture of an action as simply a natural event caused by a desire, and desire as a state of the agent external to the action, causing it and thereby (through the action, when it is successful) causing the desired state of affairs. We thereby think of the action as a certain occurrence in the world distinct from the state of affairs it is expected or intended to produce, and we think that valuing the action for its own sake must exclude valuing its end. Our vulnerability to this error is probably the reason why, when Fichte speaks of a rational constraint "to do certain external things entirely independently of external ends, merely and solely so that they happen," we think he

must be presenting us with the all but incoherent picture of an action that places no value (positive or negative) on any of its consequences, but thinks of all doing and omitting as grounded solely on some abstract description of what is done (such as “helping” or “lying”), with no thought whatever about the results. The heart of deontology, however, is that no action, properly speaking, is ever merely the effect of some desiderative state external to it. Action itself includes *rational* desire, originating in a free choice of the agent. Every action, properly speaking, must be both its own end and also, as Fichte and the existentialists see, a projection of the agent onto a future state of the agent’s world: an end to be produced which is internal to the action, rather than external to it.

6.7 Ends and Means

Now we can begin to see why although there is an important insight contained in Mill’s way of thinking about action, there is also something one-sided, incomplete, and even misleading about it, once a consequentialist gets hold of the insight and starts to twist it. Mill says that “rules of action . . . must take their whole character and color from the end to which they are subservient” (Mill, 2001, p. 2). It is false, however, to say that an action should be *subservient* to its end: this just as absurdly false as to say that reason must be the slave of the passions. It gets things exactly backwards.

The erroneous thought that an action is nothing but a means to certain ends distinct from it is what easily creates the misleading impression that a consequentialist must be the sort of dishonorable person who might do or condone anything, however wrongful or shameful, if it could be given the shabby rationalization that the end justifies the means. Mill is quite correct when he insists that a utilitarian need not be that reprehensible kind of person. There is nevertheless an important truth that lurks behind this false accusation, even though the unfair accusation misrepresents that truth. The truth in question is this: the conception of action on which utilitarianism is based is likely to miss the most basic reason why it is wrong to think about our actions as such a person does. For as soon as we conceptualize action as only a means to some end we have already represented as desirable, then it seems to follow directly that only the end could be relevant to justifying it: “The end justifies the means” thus might seem to take on the superficial appearance of a truism of universal and unqualified application. That’s why this phrase is so well suited to express a transparently sophistical way of thinking.

A consequentialist, however, need not (and should not) look at an action only in terms of the specific end to which it is most narrowly directed, as if that were the entire framework for justifying the action. A wise consequentialist can consider

the entire range of ends that must be taken into account, and reject the action that might be the best means to the immediate end because it is harmful to this range of ends taken as a whole. That is the way Mill himself would repel the false accusation. But there is something missing here even so. For the corrective that is really needed is the deontological thought that an action must always, first and foremost, be valued for its own sake, as an end in itself, never only as a means to the end (or ends) projected in it. That is the most basic reason to abstain from wrong or shameful actions, even when they are effective means to a desirable end.

An end of action is set only through a free self-projection of the agent. Therefore, rational action is always master over the end, never its slave or servant. We have reason to value the end only to the extent that we have reason to regard it, in Fichte's terms, as *formally grounded*, which means we value for its own sake the free act in which the end was projected. We have no good ground for seeking any end to be produced unless we have reason to value for its own sake the act that sets it, and therefore also value ourselves as the agent of that act. Here we see the fundamental reason why deontology gets at the deeper truth. In the sight of reason, the value of actions and agents is always more basic than the value of states of affairs as the possible goals or results of actions.

Aristotle is sometimes enlisted in support of the idea that the choice of ends does not follow under the rational faculty, on the ground that he claims that we do not deliberate about ends but only about means to them (Aristotle 1999, 1112b12–15). But this ignores the most essential end of action: the excellent deliberator, namely, always acts for the sake of the fine, which is another way of saying the action is done for its own sake (Aristotle 1999, 1116a28, b 2–3, 1117b14, 1119b16, 1122b6). It is reason, for Aristotle, that makes us aware of the good (Aristotle 1999, 1170b12), and identifies the virtuous action (the one worth doing for its own sake) with action according to right reason (*orthos logos*) (Aristotle 1999, 1103b3, 1107a1, 1115b12, 1119b18). It is always actions done for their own sake that lay down the ends distinct from them over which Aristotelian deliberation takes place, leading to decision (Aristotle 1999, 1111b5–1112b16).

Aristotle does, of course, recognize *poiesis* as a kind of human doing (*praxis* in a broad sense) that falls short of *praxis* in the narrow or proper sense. *Poiesis* thereby fails to share in the human good; at most, it contributes some of its requisites. Aristotle apparently thought of the practical arts in this way, and considered the life of the artisan to be essentially lacking because in it activity remains merely the servant of an external end. We do not have to be limited to Aristotle's ancient aristocratic view of labor and the practical arts, however. We could think of production as a form of *praxis* that is also self-affirming in its relation to the human ends it satisfies.

This is what the young Marx is doing when he speaks of people producing objects “as human beings,” affirming their human nature and their own individuality in mediating between the human species-being and the need of another individual that is satisfied by their product. “In my individual activity, Marx says, “I would have immediately confirmed and realized my true human nature” (Marx, CW 3:228). For Marx, labor is alienated when it fails to actualize the worker or the human species, but serves only as a means, producing profit for capital. It is a real question—one very important for our conception of the human good, and also a question from which consequentialist thinking tends to distract us—what kinds of production can also be valued for their own sake, hence rise to the level of *energeia* and *praxis*, and what kinds are merely the servants of desire, doomed to be mere *kinesis* and *poiesis*.

Aristotle holds not only that every activity aims at some end, but also that there is a single embracing end at which all aim in common and this encompassing end is the human good or happiness (*eudaimonia*). Virtuous action and every other good is chosen also for the sake of happiness (Aristotle 1999, 1097b1–5). It is this implicit conception of a single encompassing good that results in the impression that Aristotle commits a fallacy of composition in the first sentence of the *Nicomachean Ethics*. It is more accurate to see Aristotelian rational agency as the project of unifying goods, through a system or hierarchy of ends. Bridle-making, he says, is subordinated to horsemanship, which belongs to military strategy, which is among the crafts that seek the happiness of the *polis*. More generally, it belongs to rational agency to seek our more immediate ends for the sake of the larger and more encompassing ones, and this leads us to the idea of an end that is complete and final (Aristotle 1999, 1094a7–17).

6.8 The Encompassing Good

This idea of a single unifying good appeals to many philosophers. Mill claims that happiness is an organic whole (Mill, 2001, p. 38). Kant has a conception of a system of ends as the realm of ends and a final good as a *summum bonum* (KpV 5:110–13; KU §§ 82–5, 5:425–36). Fichte understands the ethical drive as the outcome of the I’s fundamental striving toward unity and wholeness (SL § 3, 4:39–43). That is why for him the ethical drive itself must be a mixed or synthetic drive, combining the pure drive towards rational selfhood with the natural drive towards empirical satisfaction (SL § 12, 4:147–53). He therefore claims that no pleasure, satisfaction, or enjoyment truly counts as *happiness* unless it satisfies the ethical drive, and is pursued in obedience to the moral law (Fichte, VBG 6: 299–300; SL § 18, 4:216). The right way to see this final end of action—the way we find in Aristotle, Kant, Fichte,

and even in Mill—is not to think of it as an end whose content is fixed in advance of the exercise of rational agency, but instead as the cumulative result of setting partial ends and actively systematizing them into a rational whole. This insight is the reason why Mill speaks of happiness as a “concrete whole,” composed out of qualitatively different parts (Mill, 2001, p. 38).

Sooner or later, however, consequentialist theories always seem to abandon this insight and lapse into a very different picture, in which the good is not freely determined by rational choice, but fixed in advance of free agency and rational systematization, its content constrained in advance by certain formal constraints of measurement and maximization. I think consequentialism has a tendency to do this because it is only if we so regard states of affairs, as the possible outcomes of actions, that we can represent all reasons for action as consisting solely in the value or disvalue of such states of affairs. And that is what consequentialism wants to do.

Practical deliberation, on this picture, is like shopping in a department store, as if possible future states of affairs had little price tags dangling off them, informing us not merely of the cost of making the purchase, but even more importantly of its benefits. As rational agents, we are (as bumper stickers and t-shirts put it) *born to shop*: our task is simply to pick the action that procures the best value for the money. Unlike these consequentialist agents, however, most of us who go into an actual department store have lives outside the shopping process. What has value for us once we get into the store depends more on what we went there to buy than on the observable qualities of the items sitting on the shelves, still less what we read on the little price tags. More generally, our ends provide us with the grounds for considering the properties of states of affairs as attractive or repulsive. What is best depends on rational choice, not the reverse. A phrase like “the best way for things to go” has no referent at all except insofar as rational choice has given it one in the form of a system of projected ends.

6.9 Consequentialist Reasoning Always Rests on Deontological Grounds

The ends we project, moreover, are often determined by kinds of value that don’t even pertain to the values of states of affairs considered merely as possible consequences of actions. They may be values that attach to principles of action, or to the terms of care and co-operation in which we stand to others, or to past states of affairs that provide reasons for us to act in some way or to set some end. Both actions and ends can be valued for a wide variety of reasons. And the deontological insight is that what matters most fundamentally in deciding what to do is how

these reasons converge on the question whether an action itself is good or bad, to be chosen, or avoided, for its own sake.

An artist, for example, may seek to bring into being a beautiful sculpture of Apollo or Aphrodite. But the choice to do so may be dictated less by the imagined beauty of the projected statue than by the life-choice to be a sculptor, or to be a sculptor of statues of Greek deities. The reasons for that choice might consist more in the resolution to develop and utilize a perceived talent, or the obligation to carry on a family tradition in the fine arts, or a loyalty to the sculptor's Greek ancestry, than in the desirable results, considered in themselves, of any of the possible works of art that result from pursuing this way of life.

This chapter, for example, began as a talk for a conference honoring the work of Barbara Herman. I didn't write up the talk, however, because I foresaw it as having certain attractive properties whose allure kept me working to bring them into being. Basically, I did it because a long time ago I chose the life of a philosophical scholar, which then led me into further projects, such as accepting speaking engagements, and the choice to write up my talk resulted from those choices and projects, which gave me reasons for deciding that this particular talk would be a worthwhile thing to give. That decision was more basic to my reasons for writing it up than any of the properties the talk itself turned out to have. And whether this or that feature of the talk was regarded by me as desirable, as a consequentialist sort of reason for including that feature, was more basically determined by what would contribute to making the giving of the talk a worthwhile thing to do for its own sake. More generally: which consequences of action we should treat as objects of pursuit and avoidance are determined by the ends we set. It is not true, conversely, that those ends are determined solely by the independent value of states of affairs as objects of desire and aversion, hence possible objects of pursuit or avoidance. The value of choices or actions is fundamental, and prior to the value of the resulting states of affairs.

At this point I need to warn against a misunderstanding. I do not believe, as some Kantian constructivists do, that objects in the world, or states of affairs, actual or possible, cannot have objective value properties, or that if they have such properties, they have them only if these properties are projected or conferred on them by our desires, sentiments, attitudes, or choices. I agree with Derek Parfit that a lot of our reasons are both real and object-given. The point is rather that these reasons, pointing us towards states of affairs as possible results of actions radically underdetermine the ends it is reasonable for us to set, and are therefore inadequate to determine which consequences we should pursue or avoid. Further, the objective value properties possessed by things or states of affairs often directly correspond to, and are even grounded on, the objective validity of rational principles of action.

For instance, the fact that the happiness of some person has moral value is based on the moral principle that we should make the happiness of that person one of our ends.

The principles that determine which states of affairs are valuable are, moreover, nearly always *ceteris paribus* principles: they virtually never require us to pursue some state of affairs unconditionally under all circumstances. The examples philosophers use to elicit our intuitions often conceal this point, by focusing (for instance) on the saving of innocent human lives in an emergency. But that is a rather unusual end, in that it probably does strike us as worth pursuing not only *ceteris paribus*, but unconditionally. The tacit inference from the elicited intuitions seems to be that principles covering such cases would carry over to cases where some lesser benefit is at stake. We are thus deceived into overlooking the fact that this is not true. If there are any general principles requiring certain actions or omissions, they are not likely to be justified as the only means to states of affairs we take unconditionally as ends.

6.10 “The Shadow of Hedonism”

The consequentialist view tends to presuppose, as I have said, a rather distinctive formal conception of the encompassing good (or *summum bonum*). It supposes, namely, that there is really only one end of action, which is not freely projected by us, but simply given by our nature. The consequences of actions are to be evaluated by a single standard—how far they contribute to, or fall short of, achieving this single inevitable aim. Our task as agents is reduced to determining, by empirical inquiry, which cost-and-benefit tags, as determined by this fixed end, attach to the consequences of actions, and then which actions causally result in procuring the most good and avoiding the most bad. This is the picture offered us by Bentham. The only end is happiness (pleasure and freedom from pain) (Bentham, 1948, pp. 1–2.) For any given subject, or group of subjects, we have a way of measuring the happiness that would result from the actions open to us, so that states of affairs, regarded as possible consequences of actions, may at least in principle be precisely ordered as to their desirability on a single felicific scale.

Benthamite hedonism is not as popular today as it once was. I agree with T. M. Scanlon, however, when he conjectures that even among those who do not accept it, it has cast a “shadow” over theories of the good, giving a specious plausibility to the notion that all value is located ultimately in states of affairs to be promoted, so that the good consists fundamentally in an array of future possibilities, each with a determinate positive or negative value attached to it in advance of all choice (Scanlon, 1998, pp. 100–3). But I would deny that it was really “hedonism”—the

pursuit of pleasure and avoidance of pain—that cast this shadow. Pleasure is generally pursued along with, and as part of, other goods, so that the value of a pleasure is typically derivative from (or even parasitic on) the value of something else, such as an activity we enjoy or an achievement we would value even apart from the pleasure we take in it, just *because* we already value it. That was the point of Aristotle's thesis that pleasure is the completion or perfection of activity (Aristotle 1999, 1153a2–12). Moreover, as anyone intensely devoted to a life of pleasure knows, the quest for pleasure is never part of the same project as avoiding pain. True hedonists are aware of the point Nietzsche emphasized, that we typically acquire pleasure only in conjunction with pain: “But what if pleasure and pain should be so closely connected that he who wants the greatest possible amount of the one must also have the greatest possible amount of the other—that he who wants to experience the ‘heavenly high jubilation,’ must also be ready to be ‘sorrowful unto death’?” (Nietzsche, 1974, § 12, p. 85). (Pleasure itself, Nietzsche hypothesizes, may be nothing but a certain rhythm among painful sensations, Nietzsche, 1967, p. 371.)

Perhaps consequentialist theorists do not see this because they do not have the lush, verdant souls of lovers and artists, but instead the niggardly, desiccated souls of corporate accountants or bean-counting bureaucrats. Is consequentialist foresight forever doomed to be Soames rather than Jolyon? It is if what casts the shadow is not hedonism but a set of false *a priori* prejudices about the formal structure of the good—as measurable and also fastidiously separable from the actions by which it is to be obtained. The real-world model for that good is obviously not pleasure but money. They can't talk about money, though, not only because it would be too vulgar for Forsyte 'change but also because of Aristotle's point that money is only a means and not an end. Pleasure, however, is a notion so malleable, so heterogeneous in its reference, that it can be treated as a measurable good to whatever extent our choice of ends makes it so. But the choice to make the good measurable like money is the choice of what has price over what has dignity, over objects of respect and love, over right and wrong; or else it is the choice of these too as things that might be bought and paid for. This is foolish—worse than foolish—like Soames trying to buy Irene's love.

6.11 Can Consequentialists Value Actions for their Own Sake?

The basic problem, however, is that there is no room in the consequentialist counting-house for the basic deontological truth that our choices are always to be made for their own sake, never merely for their results. There is, however, one familiar way in which consequentialists try to accommodate this point. The action

itself is considered directly pleasant or painful, desirable or undesirable, and in this way the value of the action itself is included, for the purposes of a consequentialist calculation among its own consequences. It is regarded not only as the cause of some good or evil beyond itself, but as itself part of the good or evil it produces. This is the way Bentham himself wants to accommodate beneficent actions, and also the way Mill handles valuing virtue for its own sake (Bentham, 1948, pp. 36, 142–3; Mill, 2001, pp. 36–7).

Now it may seem entirely unproblematic to think that actions valued or disvalued for their own sake can be simply folded into their own “consequences” (like shortening into cake batter) for the purposes of a consequentialist theory. Thus, an Aristotelian conception of happiness or flourishing, which consists centrally of virtuous actions valued for their own sake, might seem (as it does to Boyd) as if it is compatible with a consequentialist theory. Further, actions showing respect for persons, or for their rights, might seem capable of being given the consequentialist rationale that in them we aim at the goal of expressing respect for persons. The value of such actions might be represented as simply consisting in the way that we bring about the valuable consequence: *respect has been shown for this person*. The wrongness of violating someone’s right can be represented as a bad consequence: *a right’s having been violated*. These points may seem obvious, especially to people who want to think of themselves as consequentialists. But they are not obvious; they are not even correct. I have no doubt that by assigning the right degrees of pleasantness or goodness (or its opposite) to an action, a philosopher could manufacture a replica of whatever choice we might take to be the right one. The question is only how much this replica resembles the way we actually value or disvalue actions for themselves. I think the fake is so obvious that no Manhattan street vendor would risk displaying it.²

² This consequentialist attempt to replicate doing or refraining from an action for its own sake is only the most prominent way in which consequentialists might defend their position by reducing reasons for action of a non-consequentialist kind to consequentialist reasons—that is, reasons consisting solely in the goodness or badness of states of affairs regarded as the outcomes of actions. I think that treating Aristotelian virtuous acts done for their own sake, for instance, as having a value that can be treated consequentialistically (as belonging “trivially” to the consequences of these same acts) is equally misguided. It lies beyond the scope of this chapter to show that all such consequentialist projects do fail (still less that they must fail). However, I think it quite obvious that some obvious attempts of this kind (such as the consequentialist attempt to replicate the wrongness of friend-betrayal or other acts that are wrong in themselves) do fail and must fail. However they may be systematized by moral theories under some underlying value, reasons for action are quite heterogeneous and that reasons of many kinds cannot possibly be reduced to reasons consisting solely in the value (good or bad) of states of affairs regarded as the possible outcomes of actions. The attempt to reduce them in such a way has, I think, seemed entirely too easy to most consequentialists—so easy, in fact, that they are even shockingly blind to the basic heterogeneity of practical reasons, as well as to the fact that consequentialist reasons are not the least bit fundamental, that they themselves rest on and presuppose reasons of an altogether different kind.

Suppose the authorities offer me a sum of money to betray my friend to them, falsely implicating him in some crime. Suppose too that my acceptance of their offer will pre-empt two other people who would otherwise have betrayed their friends even more reprehensibly than I can manage with my limited talents for brazen treachery. If we provide a consequentialist replication of the deontological thought that it is wrong for me to betray my friend, investing acts of friend-betrayal with a large negative value, then we are in danger of being committed to the conclusion that I ought to betray my friend, since that would minimize the total negative value of friend-betraying acts. I have met consequentialists who seem willing to hold on to their theory in such cases, as some people did with a life-savings invested in Enron stock, all the way to moral bankruptcy. Others try to avoid this by attaching a far greater negative value (for me) to *my* act of friend-betrayal than to the friend-betraying acts of the two others—just because it is *my* act. They even have a name for this, which is supposed to make it seem more respectable: it is called an “agent-relative consequence.” Such a move, however, is not merely theoretically ad hoc but most unattractively self-centered. Why should *my* friend-betrayal, just because it is *mine*, count for more than others’? Even if we accept the conclusion that I should not betray my friend, this is no way to reach it. (For a well-argued consequentialist rejection of this dishonest ploy, see Pettit, in Baron, Pettit, and Slote, 1997, pp. 145–50.) I’ve heard this ploy defended on the ground that it enables us to unify ethical decision making with formal theories of rational choice. At that point I can only manage a wan smile, shake my head wearily and click my tongue behind my teeth: philosophers addicted to a misguided theory and trying to reduce away an inconvenient but indispensable concept will groan and sweat in any stress position, however ugly and painful, offering us grunting assurances that they find it a graceful and comfortable posture.

These options are no more appealing regarding actions to be valued for their own sake than actions to be avoided for their own sake. Suppose I am in a position to protect my friend from persecution through an act of loyalty, or suppose I value the creation of works of art as part of my vocation as an artist. It might turn out that I would maximize acts of loyalty by omitting this one, or maximize the acts of artistic creativity by spending my life as a bank-teller instead of an artist. Here too consequentialists might say either that I should omit my act of loyalty and spend my life at the teller’s window cashing checks, or else that I must value my acts of loyalty and creativity more than I do the similar acts of others. But neither of these claims is plausible.

No sensible consequentialist—by which I mean no sensible person who is looking at acts and policies from the standpoint of their consequences—will reason about the friend-betrayal case in either of these ways. Such a person will conclude

neither that I should betray my friend in order to minimize the badness of friend-betraying acts, nor that some special degree of badness is to be ascribed to these acts just because they are mine. Rather, such a person will consider that when people entertain a policy that might approve the betrayal of friends, then whatever its short-term benefits in a particular case, that way of thinking will in the long run undermine trust and friendship among people. The sensible person will conclude directly that we simply should not go there. This is how a sensible person thinks about the consequences of systematic lying, or carving up one healthy patient to save five others needing organ transplants, or any of the other examples designed to evoke those criminally insane reactions philosophers call “consequentialist intuitions.” Regarding veracity, Mill takes this sensible course:

The cultivation in ourselves of a sensitive feeling on the subject of veracity [he says] is one of the most useful, and the enfeeblement of that feeling one of the most hurtful, things to which our conduct can be instrumental; and inasmuch as any, even unintentional, deviation from truth does much toward weakening the trustworthiness of human assertion...—we feel that the violation for a present advantage of a rule of such transcendental expediency is not [after all, truly] expedient” (Mill, 2001, p. 23).

This is obvious. There is a serious difficulty lurking here for consequentialist *theorists*, however. For the very general and long-term consequences to which Mill appeals on behalf of a policy of strict veracity cannot be weighed with precision against the possible immediate benefits and harms of violating it, as consequentialist theories are so heavily invested in doing. Consequentialists often evade this difficulty in the examples they discuss by stipulating, just so as not to complicate the calculations, that the beneficial lie won’t be found out, or the consequences of its discovery will be negligible. In the real world, though, Mill is right and the tidy calculators are wrong: any *systematic policy* of deception is too likely to be found out to be a rational policy, as the Jeffrey Skillings and Bernard Madoffs of this world usually learn to their cost (and a lot of other people’s as well).

Mill gets to the right conclusion about lying, but does he get there by the best route? Reasoning from consequences is sound, as far as it goes; but it does not take into account the more basic deontological point that at least in many circumstances,³ the act of lying or betraying a friend is wrong or bad in itself, that this is the primary reason why such acts should not be done. This reason cannot be

³ It is no part of my argument here, however, that it must always be wrong to betray a friend—even if, say, the friend is engaged in extreme evil-doing and is trying to use the loyalty of friendship as a way of making you complicit in his crimes. The claim is only that the deontological thought: “You must not betray your friend” is appropriate (even decisive) in the kind of example I am imagining here, and that it would be appropriate in many other cases even where it might not be decisive. The point I am making is rather that the suggested consequentialist replication of this thought is a phony, a failure.

correctly accounted for, or even plausibly replicated, by citing some inherent “badness” among the *consequences* of these acts.

I now try to summarize. We obviously have reason to care about the consequences of our actions. But it is our ends that determine which consequences we have reason to care about. These ends are set in the course of choices and actions that should always be valued for their own sake, not merely as means to obtain results valued independently of them. No doubt the value of possible future states of affairs provides us with one kind of reason for valuing the choices and actions that set these states of affairs as ends, but there are many other kinds of reasons for valuing choices and actions, including the principles on which we should act and the kind of person we should want to be. Deontology gets at a deeper truth than consequentialism because the value of choices, actions, and agents is more fundamental than that of consequences.

6.12 Must Consequentialists and Deontologists Disagree about What to Do?

Let me end by raising a question I hinted at early on, when I said that I doubted that agents who consider consequences would have to reach different conclusions from those who approach decisions by considering reasons of other kinds. But I have to begin by noting plenty of ways in which various factors, including the imperfection of agents, the non-ideal character of their circumstances, and also the inherent limits of ethical principles and theories might make it the case that thinking about actions from the standpoint of their consequences might lead to different conclusions from thinking about them in other ways.

There could, to begin with, be significant differences of perspective between deliberations focusing on ends of action and deliberations focusing on the actions that set them. It might be better heuristically to reason from consequences in some cases, and in others from practical reasons of other kinds. But I doubt there is any systematic way to decide in which cases the heuristic advantage would lie on one side or the other. There might also be errors of perspective connected with each approach. Someone focusing on consequences might evaluate these too narrowly, by focusing on certain immediate ends and neglecting or underestimating the importance of other ends that remain in the background. Someone focusing on some non-consequentialist reason for or against action, such as a moral rule or principle, might misinterpret the rule or neglect conditions under which other reasons might justify suspending it, making an exception to it, or recognizing that an action that seems to fall under it does not really do so.

We also need to recognize that ethical theorists tend to expect too much from their principles and theories. They tend to overestimate the solubility of ethical problems (to underestimate, in other words, the absurdity of the human condition); and they also tend to think that ethical problems can always be resolved at the level of principle, when many of them are really matters of judgment about particular cases. People with the right principles, motives, ends, and priorities, if they are responsible people, will not obsess over a moral rule when it has lost its point or when consequences are at stake that are more important than it is; and if they are decent people, they will not pursue even an otherwise desirable end by wrongful or shameful means. But no theory can guarantee a right answer for every case. Some cases will be difficult, others tragic, leaving the agent no acceptable options, while even the wisest people may disagree about certain cases. Beware of any theory that recognizes only one kind of practical reason to the exclusion of all others; beware of principles that promise to tell you what you should do based on reasonings that could be carried out mechanically, even by a person of bad character or poor practical judgment.

There is a very general reason, however, based on the arguments of this chapter, for thinking that deontologists and consequentialists, if they were sensible, should ideally reach the same conclusions about what to do. An agent who evaluates consequences based on a certain set of ends, together with choices about the systematic priorities and subordination relations among these ends, should evaluate them so as to reflect the reasons that led to the choice of these ends; therefore the agent should choose those actions, and pursue those consequences, the reasons for which were bound up with the same system of ends. Whether you reason from the actions or the ends, ideally the result should be the same. *First* project the right ends, through actions that are valued deontologically—for their own sake; *then* you can decide how best to achieve those ends. Consequentialism gets left behind—not like discarded refuse or unsightly litter, but rather like the caboose of a train. Consequentialist thinking is just fine when it is doing what it is supposed to do, but it always needs the deontological locomotive in front, pulling it along the right track.

Fichte's Absolute Freedom

7.1 Fichte's Philosophical Conversion

Fichte's first publication, *Attempt at a Critique of All Revelation*, appeared rather late in his life—in spring, 1792, just before his thirtieth birthday. He had at that point been acquainted with the Kantian philosophy less than two years. Thus Fichte was in many respects already a thoroughly formed philosopher before his conversion to Kantianism, and he even thought of himself as already having a philosophical “system.” As far as we can tell, this system was largely influenced by Spinoza, and involved necessitarianism about the will. (Our best evidence of this consists in the *Aphorisms on Religion and Deism* (1790), which apparently pre-date Fichte's acquaintance with Kant, ARD, SW 5:1–8). In Fichte's own view, the pivotal point of his conversion away from this system was “the concept of absolute freedom.” As he reported in a letter of late summer, 1790 to F. A. Weissshuhn:

I have been living in a new world ever since reading the *Critique of Practical Reason*. Propositions which I thought could never be overturned have been overturned for me. Things have been proven to me which I thought never could be proven—for example, the concept of absolute freedom, the concept of duty, etc. . . . Thus I was deceived by the apparent consistency of my previous system, and thus are thousands of persons perhaps still deceived (FGA III No. 63, EW, pp. 357–8).

A letter in November of the same year to H. N. Achelis likewise describes the “revolution that has occurred in my way of thinking”:

I now believe wholeheartedly in human freedom and realize full well that duty, virtue and morality are all possible only if freedom is presupposed. I realized this truth very well before—perhaps I said as much to you—but I felt that the entire sequence of my inferences forced me to reject morality. It has, in addition, become quite obvious to me that very harmful consequences follow from the assumption that all human actions occur necessarily, and just as obvious that this is largely the source of the tremendous ethical corruption of the so-called “better classes” (FGA II, no. 70a, EW, pp. 360–1).

As we see from this very last remark, Fichte identifies his commitment to “absolute freedom” not only with the idea that duty, virtue, and morality are not illusions, but also with a *social* cause: human equality, the recognition of the rights of the poorer classes (from which Fichte, son of an emancipated serf, himself came) as against the unjust privileges of their corrupt oppressors, the “so-called ‘better classes.’” Another noteworthy claim, present in the letter to Weissshuhn, is that the denial of absolute freedom involves some kind of self-inconsistency, incoherence, or even self-deception. We will see presently that this is by no means marginal to Fichte’s conception of absolute freedom.

Absolute freedom was also to become the foundation of his Kantian philosophical system, after his rejection of the version put forward by Reinhold: “Reinhold tries to make everything that happens in the human soul into a representation (*Vorstellung*). Anyone who does this can know nothing of freedom and the practical imperative” (letter to H. Stephani, FGA III no. 171, EW, p. 371). In spring, 1795, Fichte declares to Jens Baggesen: “My system is the first system of freedom” (FGA III, no. 282a, EW, p. 385).

7.2 What is Absolute Freedom?

Fichte saw his commitment to “absolute freedom” as a version of Kantianism; at the same time, he saw his own system as the *first* “system of freedom.” We may therefore better understand the “absoluteness” of Fichte’s absolute freedom if we compare, and yet also *contrast*, Fichte’s concept of absolute freedom with Kant’s concept of transcendental freedom.

Kant and Fichte on freedom. For Kant, *transcendental* freedom is a certain kind of causal power, namely, the power of beginning a state or an occurrence (or a whole series of states and occurrences) spontaneously or “from itself” (*von selbst*) (Kant KrV A533/B561). If the human being (or the human soul) is a transcendently free cause in this sense (Kant thinks we cannot know either way whether it is), then it is a finite substance (corporeal or incorporeal—this is also unknowable), whose existence is causally dependent on other things. Substances in the natural world possess causal powers to begin a state or occurrence, but this power is itself the effect of other substances acting on it with similar causal powers—which powers are themselves the effects of other substantial causes, and so on indefinitely. The Third Antinomy of pure reason concerns the *idea* (the concept of reason, not capable of exhibition in experience) of a causal power that is not dependent in this way on other such powers, and this is what Kant means by the power to begin a state or occurrence “from itself.” Kant holds that *practical* freedom, the capacity to act without the action being necessitated by a natural impulse, requires

transcendental freedom, since any cause whose causal power results from the causal power of causes external to it (such as a natural impulse) would have its action thus necessitated (KrV A 534/B562; G 4:446; KpV 5:33). Since practical freedom is a kind of causality, and causality acts according to laws, there must also be a law for a *free* cause—the *moral law* or law of *autonomy* (G 4:446–7; KpV 5:29, 33).

Fichte accepts Kant's conception of freedom as the power to begin a state spontaneously—or, as he prefers to put it, “absolutely”; he regards the I as free in this sense (SL 4:37). But Fichte does not identify the I with a *substance* or *thing* possessing this power. Instead, Fichte conceives the I as originally only as an *acting*; it is an act that is absolutely free in the sense that it is self-positing (GWL 1:96–7; ZE 1:462–3; NR 3:1–2 n; SL 4:3–4). As a transcendental condition of the I, its action must be directed to an external object (a not-I or corporeal world); therefore, the I's free act must also be ascribed to a body or corporeal substance (the I's material body) (ZE 1:495; NR 3:56–62; SL 4:12, 110–12, 129–30). For Fichte, a disembodied or immaterial I would be impossible. “Apart from this connection with a body [the I] would not be a person at all, but would be something quite inconceivable (if one can refer to a thing that is not even conceivable as ‘something’)” (VBG 6:295; cf. NR 3:59).

Free will and the body. Here we ought to think not so much of Kant as of the more direct precursor of Fichte's doctrine: Spinoza's conception of the mind and body as the same thing conceived under two different attributes (Spinoza, 1982, *Ethics* II P7S, P13S). Or we might think of its direct successor: Schopenhauer's doctrine that the will is the body experienced directly as a thing in itself (Schopenhauer, 1958, 1:§18). But we need to keep in mind that both Spinoza and Schopenhauer think of the mind or will as a *thing* (a *being*), whereas Fichte understands it as an absolutely self-positing *act*. The body for Fichte is part of nature, and must be conceived as a living organic whole (SL 4:112–22). This natural body is not, as such, the subject of action; nor is the power of acting the I ascribes to itself a causal power belonging to the body. The active, self-positing I is rather the *will*, and the power of acting is ascribed to the I as will (NR 3:20–3; SL 4:18–23). The will and the body are regarded by the I as the same, but viewed from different sides (SL 4:11–12, 130), just as in Spinoza the body and mind are the same *res*, conceived under two different attributes: extension and thinking. But as I (or subject) I can distinguish myself even from my body, and this is because the I is *not* a thing but an act. For Fichte, the only *thing* with which I can identify myself is the body, which, however, is not the I, but rather the vehicle of the *free volitional act* that is the I. The body is a cause in the realm of bodies, acting on other bodies and being acted upon by them (GWL 1:125–36; NR 3:56–61; SL 4:11); but it is the I as will that is active, self-positing, and free in the absolute sense.

The I as will, in order to be free, must be capable of absolutely free acts, hence something that “has to exist in advance of its nature,” or to “be before it is determined” (SL 4:35–6). (Fichte never uses the famous Sartrean formula “*Existence precedes essence*,” but he is obviously the original author of the idea it expresses.¹) Fichte thus accepts Kant’s conception of transcendental freedom, but only if it is regarded as belonging to an *act* of self-positing, not if it is conceived as the causal power belonging to a thing or substance. Further, however, there is a sense in which Fichte’s “absolute” freedom is not the same as Kant’s conception of a power to begin a state “from itself.” Fichte also differs from Kant in that he does not regard the state that is begun absolutely as begun solely “from itself” in the sense that it is unconnected with something prior to it. For Fichte, to think of the I’s free act as “connected to nothing at all” would be to identify freedom with mere contingency in the sense of “blind chance,” and the indeterminacy of freedom with mere “not-determinacy=0” (SL 4:33–4, 137). Fichte consistently denies that freedom could be anything of that sort.

7.3 The “Essence” or “Original Being” of the I

Fichte agrees with Kant, however, in denying that this connection is a causal connection to some prior *being*; but absolute freedom can also not be thought of (as Spinoza does) merely as causal origination within oneself. Instead, a free act is connected to something prior—not to a prior being or causal power, but rather to a “thinking” or a “pure concept” (SL 4:35–7). The concept in question is simply that of the I’s own self-positing, its “tendency to absolute self-activity” (SL 4:37–40), or a “drive for the entire I” (SL 4:40–2), which is identified also with “the concept of [the I’s] self-sufficiency (*Selbstständigkeit*)” (SL 4:59–60). This “tendency” or “drive” is identified with the I’s “pure being,” “true essence” (SL 4:24), “essential character,” or “original being” (SL 4:29–31).

Fichte belongs to a tradition that includes Spinoza, Leibniz, and Kant, which takes it for granted that there is unconscious mental activity. The original act of the I, he often says, is unconscious, and will has its bodily origin in an unconscious striving that manifests itself in drives and impulses which must be unified or harmonized and subjected to the law of reason in the I’s endless quest for the

¹ This formula is best known from its repetition several times within a few pages early in Sartre’s popular essay “Existentialism is a Humanism” (Kaufmann, 1956, pp. 289–91); but in *Being and Nothingness*, Sartre declares: “Consciousness is a being whose existence posits its essence” (Sartre, 1956, lxii; cf. 438, and Heidegger, 1953, p. 42).

self-sufficiency or independence (*Selbständigkeit*) that is the final end for Fichtean morality.

Fichte's claim that the will is free does not involve the idea (associated with the belief in free will by some recent critics of this belief) that every experienced urge or impulse that may influence us is conscious, or that free will consists in having our behavior caused only by conscious acts. The latter thought in fact requires the defender of free will to be a Cartesian dualist, who identifies the mind with conscious states, distinct from bodily states, and equates freedom with having our behavior caused only by such states. Fichte is not a dualist, however, since like Spinoza he thinks of free acts as identical with bodily states, only some of which are *conscious* acts. The same caricature of free will also makes it easy to discredit freedom simply by appealing to the unconscious conation that has been accepted, at least since Freud (though as I've just said, its recognition goes back much farther, and this tradition clearly includes Fichte as one of its members).

Nor does Fichte hold that every phase of a complex bodily movement constituting a free action must be individually subject to immediate, conscious, rational control. This caricature of freedom has been popular in recent neuro-science-inspired rejections of free will. The free agent for Fichte is not reducible to the immediately conscious I that is, however, obviously the primary focus of free agency and the starting point of Fichte's philosophy. Rather, Fichte tries to show how starting from it, transcendental philosophy can understand the entire range of mental activity (including the pre-conscious and unconscious) and the whole complexity of intelligent behavior, as various manifestations of a freedom that is never entirely complete, but represents a striving that belongs essentially to the I as its "essence" or "original being."

It is apt to be confusing when Fichte says, on the one hand, that the I, which is *an act and not a being*, must *exist in advance of its nature* and *be before it is determined*, yet then, on the other hand, speaks of the I's having a *pure being* or *true essence*. The point of this latter talk, however, is *not* to identify the kind of nature or essence that could belong to a *thing* or *substance*—the kind of nature or essence that could figure in a causal explanation of what a thing does. Rather, Fichte is trying to characterize the *thinking* or *concept* that belongs to the I's self-positing and constitutes what is prior to its free acting. This could not be a pre-existing nature or essence, as might belong to a thing, but rather only a "necessary thinking (by the intellect) of self-sufficiency as a norm in accordance with which the intellect charges itself to determine itself freely" (SL 4:52). In other words, the concept of what the I must "be before it is determined" is *normative*. We could put it this way: *an I (or a self) is whatever must regard itself as subject to a norm requiring it to be self-active and self-sufficient (or independent) entirely to itself*. In ascribing to itself the power to act according to a normative concept, the I thinks of itself as absolutely free.

One familiar determinist argument says that if actions are not causally necessitated, then they must be merely chance or random events—which would make them not only bizarre exceptions to the laws of nature but also unintelligible as free acts. The Fichtean reply to this argument is that the alternatives “causal determination” and “blind chance” present us with a false dilemma, and that this must be a false dilemma if we are even to entertain the concept of *action*. Action occurs only if the transition from indeterminacy to determinacy is effected not by a necessitating cause but by an act of self-determination that is governed (as Fichte puts it) by “concepts” (or norms). Acting grounded on a normative concept is diametrically opposed to something’s happening “indeterministically” in the sense of its occurring merely from “blind chance” or at random. But a normative concept can be neither an *external* cause nor a *necessitating* cause: it has to be a ground of *self*-determination on the part of the agent that leaves the self-determining agent with the alternative possibilities of acting either in accord with the normative concept or contrary to it.

The norm of absolute freedom. Fichte’s concept of absolute freedom is exceedingly abstract. It may be hard to see how it is supposed to apply to morality, or even to human action in general. We may wonder, for instance, what “norm” could be constituted merely by the “concept of self-sufficiency” or “independence” (*Selbständigkeit*). The question is directly addressed by Fichte in § 18 of the *System of Ethics*, entitled “Systematic elucidation of the conditions of I-hood in their relationship to the drive for absolute self-sufficiency.” “Self-sufficiency” is identified by Fichte with wholeness or entirety, the drive for unity (SL 4:40–5; cf. VBG 6:296–9, 304–6). More specifically, it is addressed in Fichte’s argument that the true norm for self-sufficiency lies in intersubjectivity, and our relation to others (SL 4:218–30). The self-sufficiency or wholeness here is the goal of a quest, through rational communication, to unify all rational beings through an end that they can share necessarily (SL 4:230–5). This leads to Fichte’s conception of human society as church, state, and community of scholars (SL 4:235–53). In other words, the human vocation of absolute freedom turns out to be a *social* vocation: “A reciprocity (*Wechselwirkung*) through concepts, a purposeful community (*zweckmäßige Gemeinschaft*)... If all human beings could be perfect, they would be totally equal (*gleich*) to one another... The ultimate and highest goal of society is the complete unity and unanimity of all its members” (VBG 6:305–6, 310). In other words, the ideals of freedom and community, properly understood, are not in tension, but actually require each other.

For our present purposes, what is important is the way Fichte understands the conditions of human action generally as the expression of absolute freedom. Freedom is to be located in the gap between the absolute norm of reason as self-sufficiency and our own choice whether to comply with it. Fichte’s term

for this gap is “will.” Will is “the absolutely free transition from indeterminacy to determinacy, accompanied by consciousness of this transition” (SL 4:158). In other words, we are free because, and insofar as, we recognize ourselves as subject to a norm proceeding from our “pure being” or “true essence”—as an I which ought to make itself into a self-sufficient whole, and acknowledge in ourselves the power or faculty (*Vermögen*) to determine ourselves as this norm directs. According to this conception, a “will” is by its very concept *free*, and “an unfree will is an absurdity” (SL 4:159) (cf. Hegel, PR § 4, A).

7.4 The I as Will

Several things, according to Fichte, belong as such to the concept of the I as a free will. First, the I must think of itself as directed to something outside itself (a not-I or material world), towards which its power of freedom is directed (NR 3:23–9; SL 4:75–6). As we have already seen, the I is not a *thing* (whether material, spiritual, or of any other sort), but an *act*. But in being directed towards a material world, this act must belong also to something material: the body (NR 3:55–61). The I’s material body, moreover, must be an organized (living) product of nature (SL 4:101–21).

Second, it follows from this that the I must be finite and situated, limited, and also challenged by a world around it, to which the I is passive, that sets a boundary or limit to its efficacy (SL 4:97–8): “The determination to act presupposes stasis... [The I] must already have been provided with its object, and must have received it passively. Hence self-determination to act necessarily presupposes a passive state” (GEW 1:372–3). The passivity of the I is “feeling,” which is fundamental to all consciousness (GWL 1:289–322; GEW 1:365–72; SL 4:105–6). The world of the I is its own world: “My world... is determined through its opposition or contrast with me: that is, the world as I originally find it, the world that is supposed to exist without any assistance from me, is determined by its opposition to me, through its contrast with me as I necessarily find myself to be, not as I perhaps ought to make myself freely” (SL 4:72–3). Fichte is the originator of all the later existentialist ideas of “worldhood.”²

² Heidegger’s “being-in-the-world,” *Befindlichkeit*, and “facticity” are also to be understood as the transcendental conditions of all more abstract conceptions that people may form of an “objective” world (for example, the world human beings cognize through natural science) (Heidegger, 1953, pp. 52–62, 130–40). Fichte is also the author of the Heideggerian conception of spatiality as distance or “de-severance” (*Ent-fernung*) (Heidegger, 1953, pp. 101–11). Spatiality, transcendently considered, is first and foremost, that which separates us from, or alternatively, connects us to, the changes in our world that we can bring about through our free agency. “‘Object X lies at such and such a distance from me in space’ means that in traversing the space from me to the object I must first apprehend and posit such and such other objects in order to be able to posit the object in question” (SL 4:99).

Third, it follows that the I is free at all only if, and to the extent that, it *makes itself free*, by actually determining itself in accordance with the norm that constitutes its selfhood. The I “cannot ascribe to itself a power of freedom without finding in itself an actual exercise of this power, that is, an actual act of free willing” (SL 4:83). I can choose to remain passive to my feelings, failing to determine myself freely at all, and thus choose to remain *unfree*; to those who make this choice, their own freedom cannot be proven, simply because they have not *made themselves* free. “I am what I am because this is what I willed to be. I could have let the wheels of necessity carry me away. I could have let my convictions be determined by the impressions received from nature, or by the tendency of my passions and inclinations, or by the opinions my contemporaries wanted to impart to me. But this is not what I willed. I have torn myself free” (BHW 8:348; cf. SL 4:32–3). It would be pointless to try to convince someone they have some property when they don’t in fact have it. This is still pointless, even if they could acquire the property if they chose to, and if they rationalize their refusal to make this choice by claiming self-deceptively that the choice is impossible for them. Fichte does not think he can convince anyone of their own freedom unless that person chooses to act freely, thereby putting themselves in a position to be convinced (EE 1:429–35; ZE 1:508–15; SL 4:25–6).

Finally, the I “cannot ascribe to itself the power of freedom without also thinking of several actual and determinate actions as possible through its freedom” (SL 4:79). To be aware of ourselves as free is to be aware of our future as “open,” and of the “arrow” of time as fundamentally the separation of the past, which is determined as the condition of my free action, from the future, which I may determine in one or another different ways that are open to me (SL 4:97). “For us there must necessarily be a past. For only if there is a past can there be a present, and only if there is a present is consciousness possible... Consciousness is possible only if the I posits a not-I in opposition to itself. Understandably, this is possible only if the I directs its ideal activity at the not-I. This is an activity of the I and not of the not-I only insofar as it is free activity, that is, only insofar as it could be directed at any other object instead of this one” (GEW 1:409–10). Consciousness is possible, therefore, only in time, and only for an imagination that hovers (*schwebt*) between alternative possibilities (GWL 1:216–17; SL 4:136–7).

The conditional analysis of “could have done otherwise.” According to some compatibilists about freedom and determinism, to say that I could have done otherwise than I did requires no more for its truth than that it be true that if I had desired to do otherwise, then I would have done otherwise. This, they then point out, is compatible with its also being true that given my actual desires, I was causally determined to do as I have done in fact, and also compatible with my actual desires having been causally determined to be what they were. On this analysis,

then, "I could have done otherwise" counts as true even if, given my actual desires, and the determinist assumptions that I have been causally necessitated to have them and causally necessitated by them to do as I in fact do, it is also causally necessitated that I will do as I in fact do. In other words, on the conditional analysis, "I could have done otherwise" would count as true, even when it is precisely *not* true that I could have done otherwise. This is unfortunately by no means the only case where some philosophical analysis of a proposition widely accepted as true makes a mockery of Alfred Tarski's "Convention T." Emotivist and expressivist metaethics is another rich source of such cases. (For a good critical discussion of the conditional analysis, see Ginet, 1980).

Fichte's conception of absolute freedom is well suited to resisting the conditional analysis. Because the I must "be before it is," there is nothing that I am at the time of action that could be subject to causal necessitation determining what I do. Whether or not my desires might be causally determined, I could not be. Further, whatever actual desires I may have at a certain time, my formal freedom consists in the possibility of resisting them, not being determined by them in the choice of what I do. The I's "original being" is a "drive for the whole I" (SL 4:40), but the result of this drive is not to a passively experienced desire or "feeling;" it is a "thought" or normative concept (SL 4:43–5). This norm is one the I may choose to obey or choose to violate. For these reasons, it is not only true that whatever I do, I could have done otherwise, but also that at the time of any action, there are several distinct, determinate actions possible to me. It follows that the truth of "I could have done otherwise," in the only sense in which this is (or ever could be) true, is *unconditional*. It is *not* true only relative to some felt desire of mine, and it involves the presence to me of a plurality of real possibilities whose practical availability to me is not conditional on any desire I might have. Whenever I make a *choice*, it must be *unconditionally* true at that time that I could have done otherwise.

Formal and material freedom. Fichte's absolute freedom can be understood only in terms of two distinct but related conceptions, which Fichte calls "formal freedom" and "material freedom" (SL 4:135–9). Formal freedom is the power or faculty of the I to determine itself in a variety of different ways (SL 4:50–2). "Originally, that is, apart from its own contribution (*Zutun*), it is absolutely nothing; through its own doing (*Tun*), it must make itself into what it is supposed to become—This proposition is not proven, nor can it be proven. It is purely and simply up to each rational being to find himself in this manner and grant the same" (SL 4:50).

Without an act of self-determination, there is properly speaking no person, no I at all. Awareness of the I is awareness of a free action; but an act requires subjection to a norm, though not necessarily conformity to it. Such an act presupposes

both a norm of material freedom and the power of acting according to this norm, which is formal freedom. A free will must be free in a formal sense in order to be self-positing, or a transition from indeterminacy to determinacy. Material freedom is the actual determination of oneself in conformity to the norm of self-sufficiency. Only a materially free action is free in an absolute sense, but formally free action is a necessary condition for materially free action. One might try to identify “formal freedom” with Kant’s conception of “practical freedom in the negative sense,” and “material freedom” with “practical freedom in the positive sense” (or autonomy of will). Fichte does equate material freedom with the Kantian moral law, “ought” or “categorical imperative” (SL 4:55), but rejects the claim that there is a relation of reciprocal entailment between two *different* thoughts: freedom and the moral law (KpV 5:29–30). These are, he says, simply *the same thought* (SL 4:53):

In several passages, Kant derives our conviction concerning freedom from our consciousness of the moral law. This is to be understood as follows: the appearance of freedom is an immediate fact of consciousness and by no means a consequence of any other thought. And yet, as was previously pointed out, one might still wish to explain this appearance further and could thereby transform it into an illusion. There is no theoretical reason for not doing this, but there is a practical one: namely, the firm resolution to grant primacy to practical reason, to hold the moral law to be the true and ultimate determination of our essence, and not to transform it into an illusion by means of sophistical reasoning—which is certainly a possibility for the free imagination. If, however, one does not go beyond the moral law, then one also does not go beyond the appearance of freedom, which thereby becomes for us the truth, inasmuch as the proposition “I am free; freedom is the sole true being and the ground of all other being” is quite different from the proposition “I *appear* to myself to be free.” What can be derived from consciousness of the moral law, therefore, is [only] faith in the *objective validity* of this appearance. (SL 4:53–4).

7.5 The Conviction that We are Free

This difficult passage may be taken not only as explaining Fichte’s identification of our consciousness of freedom with our consciousness of the moral law, but also as an invitation to turn from Fichte’s concept of absolute freedom to his account of the grounds for our conviction (which Fichte also calls our “faith”) that we are absolutely free.

“Faith” in freedom: idealism vs dogmatism. A certain kind of popular religious apologetics goes as follows: “Both religious belief and unbelief ultimately rest on *faith*: The believer accepts the Word of God, while the unbeliever accepts the authority of reason and science. Hence the unbelieving rationalist is just as guilty as the religious person of making an irrational leap of faith. Belief and unbelief are thus on a par: The only difference is that the believer is honest enough to admit

the irrationality of his faith, whereas unbelievers try to pretend their faith is something it is not."

Those who engage in this sort of talk ought to be ashamed. Whether the believer's faith is rational or not, the acceptance of what is genuinely grounded on reason and evidence is not (could not be) irrational. The above apologetic rhetoric is nothing but a brazen attempt to represent self-deceptive irrationality as honesty and virtue, and intellectual integrity as if it were dishonesty and a vice. It is as shallow and transparent as it is disgraceful.

Fichte was, after all, trained for the Lutheran ministry, and at times he appears to present his defense of absolute freedom as if it were a version of this same absurd ploy: it appears to us, he says, that we are free, but to *dogmatists*, who think everything can be explained in terms of the causal relations among *things*, this appearance can be explained away as an illusion.

If one nevertheless decides not to explain this appearance any further and decides to consider it absolutely inexplicable, i.e. to be the truth . . . and our entire philosophy is based on precisely this decision—then this is not because of any theoretical insight but because of a practical interest. I *will* to be self-sufficient, and therefore take myself to be so. Such an assent, however, is *faith*. Our philosophy therefore begins with an item of faith, and knows that it does so. Dogmatism too, which, if it is consistent makes the claim stated above, starts with faith (in the *thing in itself*); but it usually does not know this (SL 4:25–6; cf. EE 1:429–35).

A closer look, however, reveals that Fichte is not guilty of arguing in the manner of popular religious apologetics. In the course of avowing his faith in the divine government of the world, for instance, Fichte declares that faith "should not be represented as an arbitrary assumption one may adopt or not as one pleases, as a free decision to consider true whatever the heart wishes and to do so because this is what it wishes" (GGW 8:179; IW, p. 144). Nor is the choice between "idealism" and the opponent "dogmatism" seen by Fichte as arbitrary, or a "leap" ungoverned by reasons. For one thing, this choice is rooted in the kind of person one is, and Fichte decidedly refuses to portray the two kinds of people as in any sense on a par either morally or intellectually. The idealist's faith is more honest than the dogmatist's; but this is not (as the shabby religious apologetics would have it) because the idealist admits he believes without reasons, while the dogmatist does not admit this. The idealist is more honest because his entire self-orientation is direct and open, while dogmatists are necessarily in flight from themselves and dishonest with themselves.

The starting point of idealism is the self-positing act of self-consciousness, which alone makes possible the kind of consciousness we human beings have. There is in every human consciousness some awareness of the consequences of this act, and this awareness constitutes the "appearance of freedom," which can either be accepted for the truth it is, or else subverted by being declared an illusion, and explained away as

a product of the causal action of things. To do this, however, also makes it impossible to provide an adequate account of the difference between things and our representation of them in consciousness, which is grounded in the free act of self-positing (EE 1:435–8). The dogmatist must deny this act, and refuse to accept the distinction between things and our representation of them, even though it is on the latter that our possibility of knowing things at all necessarily depends (EE 1:438–9).

By accepting as true the appearance of freedom, idealists have an immediate belief in their own worth and dignity, and a consequent commitment to morality, as well as a commitment to the search for truth for its own sake. By contrast, the dogmatist “possesses only an indirect or mediated belief in his own dispersed self, which is conveyed to him only by objects” (EE 1:433). Dogmatists are committed to saying (what they often do say), that the unavoidable appearance that we are free is an illusion.³ This admission deprives them of any immediate trust in the truth of their own convictions, along with confidence in their own worth. They cannot unite their world-view with their own self-consciousness, so they have to see themselves as the victim of self-deceptive illusions, and they’ve become used to tolerating this dishonesty in themselves whenever they can’t help noticing it. Dogmatists are, Fichte thinks, the sort of people who are at home in a social system of unjust privileges, justifying their privileges as “necessary”—if not absolutely, then at least to the existing world—as it must be, or to the status quo, represented as a social system that could claim the loyalty of people of good will. Their denial of freedom includes a cowardly, complacent unwillingness to imagine or hope for any other world or any better system.⁴ “Someone whose character is morally slack or who has been enervated and twisted by spiritual servitude, scholarly self-indulgence

³ Spinoza, for example, insists that there is in every human being an “innate” belief in their freedom of the will (*Ep.* 58; Spinoza, 1982, p. 250). Spinoza explains this belief as a result of people’s ignorance of the external causes of their actions, combined with the error of mistaking for their entire cause the partial cause of them of which they are conscious in themselves (Spinoza, 1982; *Ethics* IIP40S2a; cf. *Ep.* 58). Fichte thinks of Spinoza as the most consistent possible dogmatist (GWL 1:120). He never represents Spinoza, however, as an example of moral slackness, corruption, or self-deception, but always speaks of him with utmost respect.

⁴ “When one speaks of the technical and practical aspects of executing what is demanded by pure reason . . . the proposition, ‘we are not able to do this,’ always means the same thing. If, for example, what is demanded is a thorough improvement of the constitution of the state then the response is, ‘these proposals cannot be carried out’—meaning, of course, they cannot be carried out if the old abuses are to remain in place” (SL 4:197–8). Compare: “Those who hide from total freedom . . . who try to show that their existence is necessary . . . these I shall call *salauds*” (Kaufmann, 1956, p. 308). In the Mairé translation, *salauds* is rendered as “scum;” it has also been translated as “swine” “shits,” and “bastards,” as it also is in Lloyd Alexander’s translation of *Nausea* (Sartre, 1964, pp. 82–94). I leave as an exercise for the reader to decide which similar words should apply to the wealthy and corporations in the U.S. who sustain the Republican party in order not merely to preserve—but in the last generation even aggressively to expand—their obscene political power and economic privileges.

and vanity, will never be able to raise himself to the level of idealism" (EE 1:434). The dogmatist's stance is not confident or candid, Fichte thinks, but is characterized by passion and defensiveness: "When the dogmatist's system is attacked he is in real danger of losing his own self. Yet he is not well prepared to defend himself against such attacks, for there is something in his own inner self which agrees with his assailant. This is why he defends himself with so much vehemence and bitterness" (EE 1:434).

On Fichte's account, then, this is not a case of one irrational "leap of faith" set over against another, after the fashion of popular religious apologetics. It is highly misleading—too charitable: not to idealism, but to dogmatism—for Fichte to imply that the relation between idealism and dogmatism is something like an intellectual standoff, where "faith" must resolve a question that cannot be decided by reason. The question *is* decided by reason, and the only sense in which the dispute between the idealist and dogmatist is not decided by reason is that the dogmatist is a dishonest and self-deceptive person who cannot bear to listen to reason.

We are as yet unclear, however, on the nature of the "proof" of freedom that Fichte thinks is available to the idealist, and also why, despite this "proof," Fichte still regards faith in idealism as a "choice" made on practical grounds, rather than a clear case of detached theoretical knowledge. One thing that should be clear by now that Fichte is not basing his defense of freedom merely on the fact that we "appear" to ourselves to be free. He realizes that this fact is equally acknowledged by dogmatists, who declare it to be an illusion; Fichte concedes that the idealist cannot objectively demonstrate to the dogmatist that this appearance is not an illusion. The decisive reasons favoring idealism—those dishonestly evaded by dogmatists—do not consist in such an objective proof. Fichte does regard as one "advantage" idealism has over dogmatism "the fact that it is able to exhibit the presence within consciousness of the foundation it wishes to employ in its explanation of experience, viz. the freely acting intellect" (EE 1:430). But the advantage afforded idealism by the appearance of freedom, or the presence in consciousness of the freely acting intellect, must somehow be exploited by argument in order to show the superiority of idealism over dogmatism. These reasons rather provide a rational ground for the idealist's *decision* to treat the appearance of freedom as the truth, in contrast to the dogmatist's decision to regard this appearance as an illusion. Only in actually making this decision, subjecting ourselves to a norm of self-sufficiency, do we give ourselves reason to be convinced of our freedom. The decision in question, as Fichte presents it, is a *moral* one. The conviction of freedom depends on the act of committing ourselves to morality.

7.6 Freedom as a Moral Commitment

Don't think: "responsibility" or "blame." This commitment, however, is *not* for Fichte a commitment to freedom as a supposed condition for moral *responsibility* or *accountability*. This point is important to emphasize, because opponents of freedom of the will often depict its defenders as preoccupied with this issue—especially as seeking to rationalize what are now called the “reactive attitudes” of blame and indignation. In fact, the truth is just the reverse: it is the *determinists* who are more often morbidly obsessed with the issue of responsibility and blame. Compatibilist determinists, for example, often think they have solved the problem of freedom if they can give some plausible compatibilist account of “our” practices of holding people responsible, blaming, and punishing them, and so forth. Proponents of the “conditional analysis” usually favor it because they think it gives us “enough” freedom to justify “our” judgments of moral responsibility and blame. (They may even be right in thinking this—if “our” sick reactive attitudes are so out of control that we judge people responsible whenever they provoke our rancor or ire, without caring a whit whether they are actually free agents.) Determinists tend to be utterly oblivious to any issues regarding free agency other than those involving moral responsibility and blame.

Fichte claims that absolute freedom is necessary for “duty, virtue and morality,” but he never claims that it is required for moral *responsibility* (or even that morality requires this latter notion). Fichte's accounts of blame and punishment, which occur only in the context of right, are wholly anti-retributivist, oriented solely to the coercive prevention of acts that violate right and interfere with the external freedom of others (GNR 3:260–85). They involve no appeal whatever to freedom of the will. They could be offered just as well by any compatibilist, or even by a hard determinist who views practices of blame and punishment solely as causal mechanisms for securing external compliance with legal requirements. I submit that Fichte retained—not totally, but still to a surprising extent—the Spinozist aversion to “reactive attitudes,” even after his conversion to Kantianism. *For Fichte, the moral commitment to freedom is simply not about “moral responsibility” or the freedom we (supposedly) “need” in order to blame people or to justify our indignation at their actions.*

Fichte does argue on one occasion that the fact that we hold people accountable for an action shows we think they could have done otherwise, and therefore exhibits our commitment to regarding them as free. But the aim of this argument is very limited. It is *ad hominem* only, and addressed solely to those who might claim they have no awareness in themselves of the moral law whose content Fichte claims to be identical with material freedom. Fichte admits that some such people

might be right, if they have not made themselves subject to the law through freely positing themselves as self-sufficient. But he appeals even in their case to judgments he imagines them making about others—“[when], for example, he does not become indignant toward and infuriated with the fire that engulfs his house, but is indignant and infuriated with the person who set that fire or who was careless. Would he not be a fool to be infuriated with this person if he did not suppose that he could also have acted otherwise and that he ought to have acted otherwise?” (SL 4:62). Fichte's claim here is only that indignation towards a person presupposes that we think the person could have, and ought to have, acted otherwise. Since the argument is *ad hominem*, Fichte could concede that it cannot convince anyone of absolute freedom if (like Spinoza) the person thinks that since people cannot act otherwise than they do, fury and indignation toward them makes no more sense than fury and indignation toward lifeless objects. This is certainly not a general argument that absolute freedom must be presupposed because morality requires responsibility, blame and indignation, and these make sense only on the assumption of freedom. To my knowledge, Fichte never offers any such argument.

Moral commitment means: the actual choice of materially free action. We begin to see why Fichte really regards idealism as based on moral “faith” when we realize that he is proposing to offer morally motivated arguments for the conviction that we are free only to those who have already chosen to act on principles of moral self-sufficiency—in other words, only to those who already consciously acknowledge the kind of (Kantian or Fichtean) morality that requires absolute freedom as its presupposition. Fichte does think that the conviction of one's own absolute freedom is required for motivation to make materially free choices in obedience to one's conscience. That is the only sense in which morality presupposes absolute freedom. Morality grounds our conviction that we are free because we must ascribe absolute freedom to ourselves in order to understand ourselves as performing materially free actions in obedience to conscience and the moral law. If some people choose not to care about morality—that is, about material freedom—then as far as Fichte is concerned, those people are not in fact free, so it would be pointless to try to demonstrate to them that they are. They have *freely* chosen (in the formal sense) not to be *free* (in the material sense).

Fichte does “summon” his readers to think freely, to perform the act of self-positing of which absolute freedom consists, and then follow the path of idealism that rests upon it (EE 1:445; ZE 1:461–2). He “challenges” them to “exhibit the ethical law within [themselves]” (ZE 1:466). But he acknowledges that they may choose not to heed this summons, and if they do not choose to do so, he has no moral argument to offer them for making a different choice. No doubt Fichte himself has a low opinion (as human beings) of those who acknowledge no morality at

all, or who can accept only a version of morality that does not aspire to the heights (or, as some might prefer to say, the metaphysical extravagances) of a morality of absolute freedom; but he does not claim to have any argument to convince those people that they are absolutely free. That is why Fichte repeatedly issues the caveat that he cannot convince the confirmed dogmatist. Only those who freely choose to perform the acts he summons them to perform can come to be convinced.

A single breath from a free human being is enough to blow their system away. But we cannot refute it *for them*. We do not write, speak or teach *for* them, for there is simply no way we could accommodate them. If we nevertheless continue to talk *about* them, we do this, not for their sake, but for others, in order to warn them against the errors of the former and to divert them from such hollow and meaningless babble. Our opponents should not feel themselves demeaned by this declaration. If they do feel themselves somehow belittled by our comments, they merely reveal their own bad conscience and place themselves publicly beneath us (ZE 1:510).

Fichte argues that those who choose to be committed to morality (to a morality of absolute freedom based on a norm of self-sufficiency) are justified in the morally based conviction that they are free. This justification is twofold: first, this conviction harmonizes with the unavoidable appearance that they are free, so it does not require them to regard this basic fact of consciousness as an illusion, which would, Fichte argues, undermine all confidence in what they think they know based on conscious experience. Second, once they have made the free choice to acknowledge the moral norm of self-sufficiency, they can be made to see that their conviction that they have made this choice does not cohere with the thought that the appearance of freedom is an illusion, so that accepting this appearance as truth is the only way to think of yourself once you have made this choice. The choice to recognize the norm of self-sufficiency thus provides a further rational ground for accepting as true the appearance that we are free, a ground that goes beyond that mere appearance itself.

Freedom vs spiritual slavery. Fichte insists, however, that the idealists' conviction that they are free always remains a "faith" in the sense that there is nothing in the grounds for this conviction that dogmatists would have to acknowledge as providing *them* with a convincing reply to the thought that the appearance of freedom is only an illusion. From the dogmatist's standpoint, idealists merely choose to act in accordance with an appearance that cannot be demonstrated not to be illusory, and remain consistently committed to this illusion in their representation of their own actions. The internal coherence of the appearance of freedom and the idealist's choice to live according to a certain moral norm cannot be presented to anyone who has not already made this choice as a reason for making it, or for adopting the belief in freedom that it presupposes. Fichte thinks it cannot

be otherwise: "Any communication of conviction by means of proof presupposes that both parties agree upon at least something. How then could the Doctrine of Science communicate itself to a dogmatist, since it *simply does not agree with him upon a single point* concerning the material of cognition, and thus there exists no common ground from which they could jointly proceed?" (ZE 1:509).

We have seen that Fichte takes the appearance of freedom in our consciousness to be one advantage that the idealist has over the dogmatist. The second advantage Fichte claims that idealism has over dogmatism is that dogmatism "is quite unable to explain what it is supposed to explain, and this demonstrates its inadequacy" (EE 1:435). What dogmatism needs to explain, Fichte thinks, is consciousness itself, the possibility of our representation of the objects or things that dogmatism takes to be primary, or even exclusively real. For Fichte, that consciousness is inseparable from self-consciousness, the absolute self-positing of the I, which is in turn a free act. To deny freedom, then, is to deny a transcendental condition for the possibility of experiencing the very things the dogmatist takes to be real, and commits the dogmatist to a transcendent metaphysical system based on a "thing in itself" that can never come before consciousness. Implicitly, of course, Fichte is also claiming that idealism, by contrast, can explain what it is supposed to explain: namely, how the same world of things is grounded in our experience of them—that is, how the objective world is made possible through the absolute self-positing of the I. But Fichte never completed his own system, so from this point of view, the advantage idealism is supposed to have over dogmatism would seem to evaporate, or at least it would be deferred until someone is able to complete the system Fichte projected. As a result, Fichte tends to see the conviction of freedom as something not to be demonstrated to those who don't believe in it, but as the object of a free faith requiring a choice that most dogmatists, on account of their defects of character, will never be able to make.

Fichte's position on free will is diametrically opposed to the Spinozist position he apparently held prior to his "conversion" to Kant's philosophy in 1790. But his attitude toward the differing stances on freedom of the will, and his perception of the relationship between opposing positions on it, remains strikingly similar to the attitudes and perceptions found in Spinoza (see especially *Ethics* V P41S, P42S). Both philosophers see humanity as divided into those who act, and are free, and those who are passive, victims of their own slavish attitude of mind. Both philosophers offer the latter a path to freedom, but neither thinks many will take it, because they do not think it is easy for the free to communicate with the enslaved, nor for the enslaved even to want to escape their condition of spiritual servitude. For Fichte, this is because liberation requires a radical free choice, accomplishing an inner moral revolution; for Spinoza, it is because it requires a perfection of

the intellect as difficult and rare as it is excellent. For Fichte, as for Spinoza, those self-condemned to slavery are difficult to reach because their condition of servitude itself holds them captive in a circle of illusion, cutting them off from the truth. If we want an accurate characterization of Fichte's attitude toward those who reject absolute freedom and the morality that goes with it, I think we should sooner describe it as "contempt" or "condescension" than as "blame" or "indignation":

[The dogmatist] has always sought refuge in an appeal to some sort of original being, even if this was nothing but a crude and formless matter. But idealism does away with this completely and leaves the dogmatist standing there naked and alone. To defend himself against such an attack, the dogmatist possesses no weapons beyond the attestation of his sincere displeasure, coupled with his assurance that he simply does not understand what he is expected to do and that he neither wishes nor is able to think of what is requested. We are quite happy to believe him when he assures us of this, and in turn we simply ask that he should also believe us when we insist that we, for our part, are quite able to think of our own system. If the dogmatist finds this to be too difficult as well, then we can refrain from making even this demand and can let him think whatever he pleases about this matter, for we have solemnly confessed on many occasions that we cannot force anyone to accept our system, since the acceptance of this system is something that depends upon freedom.—As I have said, the sole recourse left to the dogmatist is simply to assure us of his own sheer incapacity, which is a purely subjective matter (ZE 1:499).

This has much in common with Spinoza's disdainful attitude toward those who are under various illusions about themselves and others, and who consequently do foolish and harmful things because they lack adequate ideas and are in bondage to irrational emotions. In this way too, Fichte always remained a Spinozist.

7.7 Freedom as a Presupposition of Theoretical Reason

Despite the apparently unbridgeable gulf between the opposed "faiths" of idealism and dogmatism, there is in Fichte one major line of argument that might be seen as having some traction with the dogmatist, even though Fichte himself does not seem disposed to think that a dogmatist might actually be convinced by such an argument. A crucial feature of Fichte's strategy in the *Foundation of the Entire Doctrine of Science* of 1794 seems to have been to begin with a transcendental account of theoretical reason, with the result "that reason cannot even be theoretical if it is not practical" (GWL 1:264). This in turn leads to

the subordination of theory to the practical; for it follows that all theoretical law are based on practical laws, or rather, since there can be only one of the latter, on one and the same law" (GWL 1:294). [Consequently,] "there is a radical extirpation of that fatalism which

rests on the assumption that our acting and willing are dependent on the system of our representations, in that it is here shown, on the contrary, that our system of representations depends on our drive and our will" (GWL 1:295). "A drive of this sort would be one which absolutely gave birth to itself, an absolute drive, a drive for drive's sake. (If expressed as a law... it is a law for law's sake, an absolute law, or the categorical imperative—*You ought absolutely.*) (GWL 1:327).

In developing this strategy, Fichte may be seen as following up a hint given by Kant in the Third Section of the *Groundwork*, when he attempts to argue that freedom must be ascribed to every rational being as a condition of its making theoretical judgments (G 4:448). Neuhouser has described in some detail the working out of this strategy in the *Foundation of the Entire Doctrine of Science* of 1794, and also explained how in the revised *Doctrine of Science* of 1797–1799, it gave way to the more direct appeal to the practical that we have just been examining (Neuhouser, 1990, pp. 41–65). Yet I think there remains something of Fichte's original intention of arguing regressively from the conditions of theoretical reason to freedom, in the form of Fichte's claim that the "proof" of freedom whose premises dogmatism lacks the power to grasp, consists in the impossibility of explaining consciousness and representation objectively, as an effect of the causality of things (EE 1:432–3, 435–40).

Understanding and conviction. Another remnant of Fichte's original strategy lies in his claim that only the idealist is capable of being convinced (*überzeugt*) of his philosophy, while the dogmatist can never be: "Spinoza could not have been convinced of his own philosophy. He could only have *thought* of it; he could not have *believed* it" (ZE 1:513). Conviction (*Überzeugung*), according to Fichte, always arises out of a condition of doubt, a condition of worry or concern (*Besorglichkeit*) in which "the imagination continues to waver between opposites" and ending with a "feeling of harmony" or "satisfaction," through which this wavering ceases, and "the power of the imagination is now bound or compelled, as it is in the case of everything real" (SL 4:167). This "wavering of the imagination," according to Fichte, requires freedom, just as much as willing does (GWL 1:238–9; EE 1:423; SL 4:67–8, 136). In this respect, the acquisition of "conviction" has freedom as its precondition in the same way as the acquisition of "understanding" (*Verstehen*) or "comprehension" (*Begreifen*):

First of all—what does it mean to *understand* or *comprehend*? It means to *posit as fixed*, to *determine*, to *delimit*. I have comprehended an appearance if, through it, I have attained a complete cognitive whole that, with respect to all its parts, is grounded in itself; i.e. if each part is grounded or explained through all the others, and *vice versa*. Only in this way is it completed or delimited.—I have not comprehended something if I am still in the midst of explaining it, if my interpretation of it is still in a *state of wavering* (*Schweben*) and therefore

not yet fixed; i.e. if I am still being led from one part of my cognition to the others. (I have not yet comprehended some contingent A, if I have not thought of a cause for A, and this means—since a particular kind of contingency must belong to A—if I have not thought of a particular cause for it) (NR 3:77).

Here Fichte contrasts the state of *having understood or comprehended* with a temporally preceding state, in which one was still *coming to understand or to comprehend*. This preceding state is a hovering over or wavering (*Schweben*) between alternative possible judgments about what causes the object still to be comprehended, or (more generally) about how a cognitive whole relates to its constituent parts. Fichte seems to assume that for any fact to be theoretically understood or comprehended, there is a determinate cause, or grounding relation, which, when comprehended, will remove our sense of contingency from the object. Here the modal conceptions one might use in describing this contingency are merely *epistemic* modalities. For instance, first I have the thought: "It is *possible* that A is caused by X, but also possible that it is caused by Y." Then I come to think: "No: I see now that X *could not* have caused it, so it *must* have been caused by Y"—resulting in my comprehension of A as caused by Y. In this way, the possibilities regarding the object over which I hover, or between which I waver, are only epistemic possibilities: *for all I know*, A could have been caused by X or by Y. But once we comprehend A, it becomes *certain* what caused it, and so I can no longer judge otherwise than that Y caused it. My *judgment* then becomes *epistemically* necessary, even if the fact in question of what caused A is only a contingent fact. Dogmatism can perfectly well account for the fact that prior to arriving at understanding, comprehension or conviction, the mind "wavers" between these epistemic possibilities, regarding them as epistemically contingent. For the dogmatist, it can be causally necessitated that a given person remains confused or ignorant, hence in a state of epistemic contingency (or uncertainty) about many things. (This is in fact precisely the account Spinoza gives of this phenomenon at *Ethics* IP33S1 and IIP44CS.)

The temporality of understanding. Yet there remains one crucial aspect of the wavering of imagination that cannot be accounted for in the same way: namely, the wavering of my mind itself, of which I am conscious prior to the act of comprehension, between the two alternative epistemic possibilities. In order to come to comprehension, I must entertain the two epistemic possibilities as possibilities (of some future fixing, determining or delimiting judgment) that are at this point in time still contingent in the sense that I must regard them as *open to me to judge*. Here I cannot regard the possibilities as merely epistemic, due only to my ignorance. I cannot regard my considering it contingent which of them I judge merely as a matter of my being ignorant of what my eventual judgment was already

determined to be. The temporal process of coming to comprehend is possible only through a wavering at time t_1 between alternative possible judgments and then my coming to fix or settle on one of them at a later time t_2 . To regard the process as a genuine case of my coming to comprehend requires that both were really open to me to judge at t_1 , in order that I may settle the matter by achieving comprehension at t_2 . If instead I regard the matter as having already been settled or necessitated beforehand (unbeknownst to me) at t_1 , that amounts to a denial that my judgment at t_2 came about precisely through the process of my (first) wavering between opposed epistemic possibilities and (then) settling, fixing, determining, or delimiting the matter to one of them. Even where the possibilities are merely epistemic with reference to the object (for instance, the possible causes of A), the *possibilities of judging or fixing* that are open to me at t_1 must be more than epistemic. They must be real contingencies, capable of being settled by me, and at t_1 still open to my free agency to fix and determine later at t_2 . This freedom, moreover, must be *absolute* in Fichte's sense: I must *be indeterminately* as questioner before I am later (self-) *determined* as judge. Or in Sartrean terms: in order to judge at a time I must put myself *outside of being*: my *existence* as questioner must *precede my essence* as judge.

Sartre in fact closely follows Fichte in arguing that the question is the proof of freedom: "The questioner, by the very fact that he is questioning, posits himself as in a state of indetermination; does not know whether the reply will be affirmative or negative" (Sartre, 1956, p. 5). "It is essential that the questioner have the permanent possibility of dissociating himself from the causal series which constitutes being and which can produce only being . . . He must be able to put himself outside of being" (Sartre, 1956, pp. 23–4).

Fichte thinks the process of transforming the wavering of imagination into the fixity of understanding is essential at every moment of time to our coming to the justified conviction of the reality of the material world around us. "Intuition," he says, is "fixed or stabilized by reason," so that an object can be considered one and the same in different determinations of it. Imagination then "wavers between conflicting directions," then through understanding "the transiency [of this wavering of imagination] is arrested, settled, as it were, or brought to a stand, and is rightly called *understanding*" (SW 1:232–3). Fichte cites certain philosophers (Salomon Maimon would seem to be among them) who have come to realize that ordinary understanding is a result of the exercise of imagination, but have been tempted by this to consider the entire process a deception (SW 1:227, 234). Fichte insists that they must be mistaken; for that which I must represent as necessary to successfully coming to understand cannot be represented as deception. That would make understanding itself a deception, which would be nonsense.

To be sure, it is always possible that some particular judgment (about the cause of A, or about the reality and the determinations of some object before me) was indeed pre-determined beforehand, by a process I did not understand, so that it did not after all come about by means of the wavering of my imagination and its fixation by understanding. But this is to exhibit this particular supposed act of coming to understand as a deception. If I represent my judgment as pre-determined, then I cannot, on pain of incoherence, also represent it as a genuine case of understanding or comprehension. This remains the case even if it resulted (even non-accidentally) in a *true* judgment about the cause of A. This is because it is a conceptual point about comprehension that the process of coming to comprehend must be essentially self-transparent: if it came about in such a way that the subject is essentially deceived about how it came about, then *understanding* or *comprehension* has not come about at all. A subject need not, of course, be conscious of every aspect of the process—for instance, of all the neuron-firings—that went into it. But the subject cannot be essentially self-opaque or in error about the normative-epistemic essentials of the process. And these include the real (not merely epistemic) contingency of the possibilities of judging that are open to me at t_1 . If I were mistaken in believing at t_1 that my judging was genuinely contingent, dependent on the course of my thinking—that it was truly open to me at t_1 make any of several really possible judgments about the cause of A—then the judgment I do eventually make at t_2 about this cause could not possibly be a genuine case of my coming to understand or comprehend A.

Freedom and reason. Another way to think of it is this: to come to understand or comprehend is to come to judge something for a *good reason*, and to have one's judgment determined by that reason. A reason, however, has the peculiar property that although it may explain why I judge as I do, it does not do this by preventing me from judging otherwise or taking away from me the genuine possibility of judging otherwise. So a judgment made for reasons is always contingent, not merely epistemically, but really contingent. Reasons, in other words, even the best reasons, always leave us free to act against them. (Reasons, as Leibniz said, “incline without necessitating.”) This is why the conspicuous truth that humans are the only irrational animals is grounded on the deeper truth that humans are the only rational animals.

Or again, we can also see it as a consequence of a point made earlier about acting on norms. Subjection to norms is just as essential to theoretical judgment or assent, hence to coming to understand or comprehend or be convinced, as it is to any other action. But it is only free acts that can be explained by their conformity to norms. To be aware of a norm as governing one's action, and to regard oneself as required to act (including: to judge) according to the norm, is therefore to

presuppose that one is free in this acting (or judging): that the action (or judgment) is one of several really (not merely epistemically) possible acts (or judgments) and that it depends on me which of these possibilities I settle on or determine.

This is not like a case where something might appear to me to be one way, where I can know it is really another—as when the sun appears to move across the sky, but I know that it is really the earth turning on its axis, or the Müller-Lyer lines appear unequal but I know they are really equal. In these cases, I am not rationally committed to assent to the “appearance,” as I am rationally committed to assent to the claim that my understanding or conviction came about through a free judgment for reasons whenever I represent myself as coming to understand or be convinced. To suppose my acts of judging, understanding, comprehending, coming to be convinced for reasons, were all pre-determined beforehand is thus in effect to suppose that all such acts are illusory and self-deceptive. Unless I think of myself as having the capacity to judge for reasons between genuine possibilities, I must view the entirety of my epistemic life as self-opaque, a deception perpetrated on me (though necessarily with my own complicity) by external causes of which I am unaware.

Philosophers are routinely accused of overestimating the extent to which human beings act rationally. In view of the role irrationality (evil, weakness of will, sophisticated rationalization, self-deception, self-opacity, and the like) plays in the moral psychology of rationalist philosophers such as Descartes, Kant, and Spinoza, this charge seems as silly as it is ubiquitous. Some philosophers even like to toy with the thought that all our conscious processes involve self-opacity or deception. Nietzsche, post-modernists, and neurophilosophers, for instance, sometimes enjoy titillating themselves with such thoughts. In dreams, Nietzsche says, we often imagine a cause for a particular sensation (say, a far-off cannon shot we hear). The cannon shot appears to us later in the narrative of the dream than its supposed cause, but in fact the representations produced by the state have been misunderstood as its causes. “In fact, we do the same thing when awake . . . We want to have a reason for feeling this way or that, [and] we admit this fact only—become conscious of it only—when we have furnished some kind of motivation . . . Thus originates a habitual acceptance of a particular causal interpretation, which, as a matter of fact, inhibits any investigation into the real cause—even precludes it” (Kaufmann (ed.), 1954, pp. 496–7). Or, as a more recent neurophilosopher puts it: “Your conscious life is nothing but an elaborate post-hoc rationalization of things you really do for other reasons” (Ramachandran, 2004, p. 1.)

The only point a philosopher like Fichte wants (or needs) to make here is that there is a *limit in principle* (on pain of incoherence) to the extent to which we can coherently represent *all* human thinking as self-opaque, deceptive, or based on

post-hoc rationalization. I cannot, for instance, represent my belief that all human thinking is self-opaque as a rational judgment to that effect if I include that very judgment within the scope of human thinking that suffers from self-opacity. Nietzsche, or a neurophilosopher, cannot suppose that the same deflationary accounts they give of the mental lives of others' life apply to the acts of inquiry and discovery through which they have concluded that people act for reasons other than the ones of which they are conscious. To suppose this would be to discredit their own accounts of other people's mental lives.

How Spinoza avoids the incoherence. Spinoza does not belong to this species of philosophical incoherence, though he may face a different kind of problem. Spinoza holds, regarding certain privileged epistemic states, which he calls "reason" and "intuition," that they involve comprehension of their own necessity (Spinoza, 1982, *Ethics* IIP4oS, pp. 41–4). He dismisses as illusory all states involving "imagination," which (he agrees with Fichte) necessarily wavers between alternatives, representing its objects as contingent (Spinoza, 1982, *Ethics* IIP44CS). Spinoza might be quite correct in cases of understanding that have been settled in the past. Once I have seen clearly and distinctly that $2 + 3 = 5$, then as long as I retain the results of this insight as part of my beliefs, it will no longer be possible for me to believe that $2 + 3$ could equal any number other than 5. It is also true of many of our beliefs that they were not arrived at directly by such a process, but were acquired along with an entire web of beliefs which was arrived at through temporal processes yet without each of them being acquired separately by such a process. Spinoza's position, therefore, might be coherent if reason and intuition are considered atemporally, as states in which the subject has always understood something, as it were, without ever needing to come to understand it. For then they might be necessitated in the same way as our standing conviction that $2 + 3 = 5$.

Spinoza also holds that time itself is a product of imagination (1982, *Ethics* IIP44CS). Perhaps this implies that the temporal process of coming to understand presupposes the erroneous affirmation of real contingency, which is seen through once the object has been fixed through comprehension. Fichte's argument, however, proceeds from the premise that the temporality of our understanding is a transcendently necessary condition of our experience of whatever we can ever understand, and is consequently real and undeniable. Nothing, in other words, has ever been understood except through a temporal process of coming to understand; freedom is a presupposition of this temporal process; and this is what Spinozist necessitarianism cannot account for: "All our consciousness commences with indeterminacy, for it commences with the power of the imagination, which is a hovering (*schwebendes*) power wavering (*schwankendes*) between opposites" (SL 4:194). It is essential to our lives as temporal experiencers and knowers that we are

confronted at every moment with the task of achieving some new understanding that presupposes first a wavering between alternatives and then a fixation on one of them.

Thus instead of saying that Spinoza could never be convinced of his philosophy, Fichte might better have said that Spinoza could never have *come to be convinced* of it—adding, however, that *coming to be convinced* is, for a human subject, the only way of ever getting into the state of *being convinced*. Fichte might say that dogmatists might be convinced of their system, or comprehend the world in terms of it, if only they could accomplish this without ever having to *come to be convinced* or to comprehend.

Perhaps this is how Spinoza's eternal God understands things, but it cannot apply to existing human beings, who are always in the process of becoming—except that Spinoza's God has no understanding (*intellectus*) at all, or if God can be said to have an understanding, it is no more like ours than *Canis Major* up in the night sky is like the little animal barking by our feet (1982, *Ethics* IP17C2S, cf. IP31). Or perhaps it is the way Spinoza thinks we know things by the “second kind of knowledge” or “reason”—which is apparently the way the *Ethics* itself is being represented *more geometrico*. However, when Spinoza thinks of himself as knowing by the “third kind of knowledge” or “intuition,” he supposes a spatio-temporally situated kind of knowing, where he directly perceives that the essence of a certain thing (for instance, of one of his own mental states) follows by common notions from an adequate idea of certain attributes of God (IIP40S2). Spinoza also seems to think, however, that this third kind of knowledge can be achieved *sub specie aeternitatis*, so that what is known is not represented as having come to be intuited by an individual mind at a specific time (VP22, P29). In his unfinished treatise *On the Improvement of the Intellect*, Spinoza reports that “the things I have been able to know by this kind of knowledge are as yet very few” (TIS, p. 238). We might regard even this as a wild exaggeration.

“Externalist justification.” A notion of understanding without coming to understand might be what some epistemologists are trying to achieve in another way by adopting exclusively “externalist” theories of justification. These might be used to show how subjects can be justified in holding certain beliefs (hence perhaps in understanding or knowing them to be true), viewed entirely from an external perspective (hence perhaps not available to the knowing subject itself). Dogmatists might then be justified, in this externalist sense, in holding a view about themselves that could not be coherently combined with any account they could be in a position to give of how they arrived at this view, or came to be convinced of it, or came to think of themselves as justified in holding it. The problem with this, however, is that the judgments of the externalist epistemologists

themselves—their judgments about how people's convictions are justified—could in turn be coherently declared justified only by others, and never by these epistemologists themselves. We thus are offered a picture according to which people might have all kinds of justified convictions, or even scientific knowledge, but no one could be in a position to say that they do, based on reasons they can claim as part of their own intellectual life. This is a picture it would be in principle impossible for you rationally to accept based on reasons or evidence available to you (since that would have to be “internal” justification). And that's a sufficient reason for any human being to reject this picture, even if it cannot be shown to be objectively false.

“The first person standpoint.” These reflections might provoke the objection that this argument could be valid only from a “first person” standpoint—which might then be depicted as “subjective” in a way that would allow us to dismiss it as illusory. But, of course, the philosophers or scientists who engage in this act of dismissal would need to adopt a first person stance on their own acts of this kind, and they could not then coherently demote *those* thoughts to the status of subjective illusions. In any case, it is simply not true that the presupposition of freedom involved in the attribution of understanding, comprehension, and conviction applies only to a first-person stance on our own mental lives. For we must equally represent any other person as free in this way if we propose to offer them reasons for any conviction, and coherently suppose that they might be convinced by these reasons. By the same token, if we want to think, in the third person, of still others as capable of being addressed in this way, and of coming to be convinced for reasons, we must presuppose that they are free as well. There is no “purely objective” stance we can take on people that regards them as lacking freedom, unless we also cease to regard them as beings who can self-consciously *act*, or *interact* with others, or even (with comprehension) expect to *observe* others acting *for reasons*. Thus Gazzaniga, 2011 (pp. 143–79), might be on to something (and in agreement with Fichte) when he argues that freedom is a property of the way people interact. But he reverts to tedious compatibilist orthodoxy (and departs from Fichte) in seeing the issue solely as about “our social practices” of legal responsibility, blame, and punishment. For Fichte, I-hood and rationality themselves belong only to beings who are “summoned” by others, so that a rational being can exist only insofar as there are many rational beings in communication with one another (NR 3:17–56; SL 4:218–25).

Take the issue of free will itself: suppose I am trying to decide what position is correct on this issue—Fichtean libertarianism, or some necessitarian position, whether Spinozist or soft-determinist. No matter how good the arguments on the necessitarian side of these disputes may be, the arguments always arrive on

the scene too late. In order even to entertain them as rational arguments, I must already represent myself as having a variety of possible judgments open to me, in order to be capable of deciding the question at a time and according to reasons—in other words, I must already presuppose that necessitarianism is false. If I represent myself as coming to judge for reasons that my judgment is necessitated, then I thereby commit myself to the position that the judgment in question could not have been a judgment for reasons after all. *For me*, that has to invalidate the necessitarian arguments (whatever their apparent merits) as possible rational grounds for my conviction. As Derek Parfit points out, radical skepticism about epistemic norms, although self-defeating in the way I have just pointed out, has been held by many recent philosophers, among them, Quine and the later Wittgenstein (Parfit, 2011, V. 2, pp. 521–5). Such positions are content, for instance, to provide psychological accounts of our epistemic practices, while rejecting the very notion of a “normative epistemology” that might provide grounding even for these psychological accounts.

In short: the necessitarian position cannot be coherently combined with the thought that I have come to be convinced of it for good reasons. But exactly the same goes for my attempt to represent *you* as having adopted necessitarianism about yourself for good reasons, or *our* attempt to represent *some third person* as having done this. The problem is not that we take a first-person standpoint when we consider human agents (though of course we must do that, just as we must adopt the first person standpoint when we consider *anything*), but rather that no one can ever coherently represent any conscious agent (any being that can and must take a first-person standpoint and act for reasons) merely as a causally necessitated mechanism. Fichte sees this point clearly:

The relationship between free beings is one of free interaction; it is by no means a relationship of mere causality operating through mechanical forces. . . . [In seeking to convince others] we begin with freedom, . . . and assume that they are free as well. To be sure, in presupposing the thoroughgoing validity of the mechanism of cause and effect, [the dogmatists] contradict themselves. What they say stands in contradiction to what they do; for to the extent that they *presuppose* mechanism, they at the same time elevate themselves above it. Their own act of thinking of this relationship is an act that lies outside the realm of mechanical determinism. Mechanism cannot grasp itself, precisely because it is mechanism. Only a free consciousness is able to grasp itself (ZE 1:509–10).

Why the conviction of freedom must remain a “faith.” It is important to see, however, that the considerations just adduced do not show directly *that I am free*, or that my judgment ever actually selects, for good reasons, between genuinely contingent possibilities for judging. The above Fichtean arguments leave it still possible, considered abstractly in itself, that we never really come to understand

or to judge for reasons at all, that all our supposed comprehension or understanding is always illusory. They show only that it belongs to the concept of coming to understand that it must involve freedom, and hence that we can never coherently represent ourselves—to ourselves, or to others—as coming to understand that we are not free (or, indeed as coming to understand anything at all, unless we are free). This is the most basic reason why Fichte holds that we can never finally prove that we are free, or that idealism can ultimately refute dogmatism, so that the idealist position must be described as based on *faith*.

But this way of putting it, however, may also be misleading. The “faith” in question is not the least bit arbitrary, irrational, voluntary, or even avoidable as long as we are thinking coherently about the world along with our own thoughts concerning it. There is no alternative to the conviction that we are free as long as we represent ourselves as understanding or comprehending anything, or having come to be convinced of anything at all through reasons. The necessary representation of ourselves as free involved in such self-representation is not a *psychological* necessity, something we “can’t help thinking is so” (but from which a less frail or defective mind might be exempted). It is rather a *normative* necessity, arising as soon as I try coherently to combine my claim that I have come to understand with any representation of the process through which I have come to understand. Freedom is not so much proven as presupposed by all doubting, questioning, coming to be convinced, as a necessary condition of their occurrence. Compare “Anguish is not appeared as a proof of human freedom; the latter was given to us as the necessary condition for the question” (Sartre, 1956, p. 33).

Assent to the presupposition that we are free is required if we are coherently to represent ourselves or someone else as having engaged in these cognitive acts. Fichte observes that the task of deciding between alternatives for reasons is continually our task as knowers or intelligences, simply because it is more fundamentally our task at every moment as agents—that is, as *willing* beings. For whenever we find ourselves in reflection, Fichte argues, we always find ourselves fundamentally as *will* (SL 4:18–23). It is the *willing* I or the “practical I” that is always the original I of self-consciousness (NR 3:20–3). And volition, along with the process of coming to understand or being convinced, is something we can coherently represent only as a conscious transition from indeterminacy to determinacy (SL 4:137; cf. SL 4:79).

7.8 The Traditional Problem of Free Will

The traditional problem of free will, whatever position one may take on it, is the problem of reconciling human agency with our metaphysical conception of the

world and how things work in it. Epicurus, who made the startling discovery of this problem around the beginning of the third century B.C., tried to solve it by the desperate act of postulating a “swerve” to the motion of the atoms that would allow for contingency, rationality, and accountability, and rescue us from the threat of fatalistic necessity. The reconciliation of freedom with our view of the objective world is equally the problem for naturalistic incompatibilists, such as Epicurus himself, and naturalistic compatibilists, who try to conceive of freedom in such a way that it can be fit more easily into the causal order of nature. The problem is not essentially different, however, from that of those incompatibilists who want to locate freedom among the faculties of Cartesian immaterial thinking substances or supernatural noumenal selves. For even their anti-naturalist picture is an attempt to make freedom *compatible* with our view of the world as a whole and the nature of the things or entities it contains.

We might be tempted to think Fichte belongs among those who seek such a supernaturalist solution. He does occasionally assert that the failure of dogmatism and its “materialistic” conception of the world justifies us in claiming that as free agents we belong to an intelligible world (SL 4:91; GGW 8:181). But we misunderstand such remarks if we take them to refer to an attempt to explain freedom in terms of a theoretical metaphysics of noumena or things-in-themselves (such as Kant is often represented as doing). Fichte’s own conception of the “supersensible” is part of his philosophy of religion, whose status in relation to his ethics and transcendental philosophy is by no means obvious. But to take his claim that we are members of an intelligible world as asserting a transcendent metaphysics of supernatural entities cannot be correct, for this would be to transform Fichte’s philosophy into a species of dogmatism. When Fichte claims that dogmatism is incapable of explaining the I in terms of “things,” he means to exclude spiritual or supersensible as well as sensible “things.” (That this is not really Kant’s position, and in any case could not be a self-consistent Kantian position on freedom, is argued in Chapter 7 of Wood, 2008). For Fichte, the I is necessarily embodied in a natural, living body: the very idea of a non-physical, spiritual entity as the subject either of our thoughts or our free actions is incoherent and inconceivable. Whatever Fichte may be, he is not an incompatibilist-indeterminist who proposes to explain free action through a supernaturalist metaphysics postulating supersensible things-in-themselves.

Another possible stance on the traditional problem of free will (or indeed, on any philosophical problem), is not to solve it (either naturalistically or supernaturalistically), but rather to declare it *insoluble*. Philosophical problems can be treated this way either by being dismissed on some pretext or other as pseudo-problems, or by acknowledging them as permanent sources of dissatisfaction and perplexity

that we have to learn to live with, however frustrating and humiliating this may be for us, as part of our human condition. The latter stance seems to be Kant's final word on the problems he explored in the Transcendental Dialectic (KrV A vii–viii, A293–8/B349–55, A338–40/B396–8). When he rejects dogmatism as a system of philosophy, Fichte is in effect taking this Kantian position with regard to the traditional problem of free will. The only way freedom can (or needs to) be reconciled with our view of things is that its acceptance is a transcendental condition of the possibility of being conscious of an objective world, representing objects, or coming to convictions about them, or understanding or comprehending them. For the Fichtean idealist, freedom falls outside the objective world—not metaphysically, as if it belonged to some *other* world (for no other world would be conceivable for the I), but rather methodologically or transcendentially. Fichte's position then is that the *insolubility* of the traditional problem of free will is a transcendental condition of the possibility not only of morality and agency, but even of our conscious experience of any objective world.

Fichte expresses the point this way: “Freedom is our vehicle for cognizing objects, but the cognition of objects is not, in turn, the vehicle for cognizing our freedom” (SL 4:79). Thus in order to make the problem of free will soluble, we must reject or leave behind the traditional problem—that of somehow fitting freedom into our conception of a world of objects that are merely present at hand independently of us. Instead, we must begin with the assumption of freedom, and conceive our world of objects as something available for modification through freedom. In thinking of the world as an object of natural science or theoretical inquiry, for instance, we must think of it as a place in which we can choose between alternative ways of inquiring, different ways, for example, of designing and performing experiments, making observations, and drawing inferences from them.

A natural world from which our free agency has been excluded, or into which we hope later to integrate it by reinterpreting freedom compatibilistically as the operation some sort of necessitarian causal mechanism, is a world that we could also never know, precisely because it is one in which we could never freely act. The problem of freedom is soluble only when approached differently, in a critical or transcendental way. “Our world,” as Fichte puts it, “would be originally determined through freedom as a theoretical principle” (SL 4:74). This is a world already characterized by the distinction between what is necessary and what is contingent for us, a world in which our freedom has been assumed as a condition of the possibility of acquiring knowledge about the world. This is, of course, the world in which all human agents, including natural scientists, already live, and from which naturalistic theories could provide an escape only by excluding them from the knowable world, as knowers as well as agents.

8

Fichte's Intersubjective I

8.1 Our Cartesian Habit

As philosophers of mind, we are all recovering Cartesians—in the same sense that some people are said to be recovering alcoholics. Like recovering alcoholics, we have our good days and our bad days. I sometimes think that meetings of philosophers of mind ought to begin by going around the room with introductions saying “My name is—” —and here too it might lead to greater candor if we confined ourselves to first names only—“and I am a Cartesian.” The state of denial involved in pretending otherwise seems to me what condemns most philosophy of mind to failure from the start.

Cartesianism is by no means only a problem for those few who still believe in an immaterial non-extended substance as the subject of thoughts. The deepest and most intractable Cartesianism belongs rather to those who think that thinking, consciousness, action, the mental itself—belongs not to the whole human being, embodied and situated in a larger natural and especially a social world, but rather to something “inside” the whole, situated human being: to only a part of it—for instance, to neural processes going on inside the brain. This includes all those who regard human thoughts and feelings as “supervenient” on brain states, and think we are investigating the mental when we do brain science, or even inspecting the mind’s activities when we examine MRI images. This conception is much closer to Descartes’ original conception of mind than people usually realize, since Descartes was in fact the first great philosopher to try to relate thoughts to brain states. The fact that he thought the relation had to be causal—a relation between states of two distinct substances—is far less central to what was really wrong with Cartesianism than his even more basic assumption that thoughts are a function of something that must be accessible to physics or physiology, rather than something that can be properly understood only through human interactions in a social and historical world. Truly to overcome Cartesianism in the philosophy of mind, one

would have to overcome a lot of what these days passes for “scientific” philosophy of mind. We are not even close to doing this as yet in our philosophies and sciences of the mind. To suppose that one has corrected Descartes merely by denying the substance dualism of his solution is to fall into the same state of disastrous denial as the alcoholic who tells himself he is only a “social drinker” who has his “problem” well under control.

The real struggle against Cartesianism began much earlier in the continental than in the Anglophone tradition, because the latter has always been dominated by empiricism. Empiricism consists in a set of mistaken a priori prejudices about what experience is, and how we learn from it, which has worked a lot of mischief both in modern philosophy and in the sciences that grew out of it. Empiricism emerged in the middle of the eighteenth century from a creative misreading of the modestly skeptical Cartesian rationalist, John Locke. Locke in turn took over rather uncritically Descartes' conception of mind and experience, even if he accepted it in a skeptical spirit foreign both to Descartes and to the ambitious materialistic versions of Cartesianism that have prevailed since the eighteenth century. My aim in this chapter is to explore some of the more genuinely anti-Cartesian thoughts that have emerged in the nineteenth- and twentieth-century continental tradition, especially the thoughts of the philosopher I regard as the earliest decisive anti-Cartesian, Johann Gottlieb Fichte.

8.2 “Descartes' Error”

In the modern counter-movement to Cartesianism, however, whether continental or Anglophone, there has been little agreement among philosophers about how Cartesianism is to be avoided. More often it is simply *evaded*; the anti-Cartesian philosophical schools are always more successful at labeling their rivals as “Cartesian,” or “solipsistic,” or “monological” than they are at developing a clear and convincing alternative.

Thus, we have to read far into Antonio Damasio's *Descartes' Error* (Damasio, 1994), before finding out what Damasio thinks Descartes' error was, and then it is not clear that he locates any view at all that is both clearly an error and was clearly held by Descartes, except the metaphysical dualism that is now so rare even among philosophers that it seems silly to waste the title of a book on it. Apparently the chief thing Damasio wants to criticize as “Descartes' error” is Descartes' alleged partitioning of intellect or reason from emotions and their rootedness in the body. But this is based on a (common, but nonetheless egregious) error about what Descartes' actually thought on this question, perhaps due to ignorance (or misreading) of his late treatise *The Passions of the Soul*. The misreading is typically

occasioned by the drawing of (fallacious) inferences from the conjunction of Descartes' metaphysical dualism and some of the strictures of his philosophical method to very general conclusions on his behalf about how the human intellect works best. In fact, Descartes recognized the influence of the body on the mind, and especially emphasized the role of the passion he calls "generosity" in helping the intellect to function well in the general affairs of life. It is true that Descartes did not explicitly formulate what Damasio calls the "somatic marker hypothesis" in his account of the intellect, but he certainly anticipated Damasio's general idea when he recognized that certain passions (which for him are motions of the animal spirits in the brain), help the intellect to keep other passions in check and constitute a kind of bodily infrastructure for the intellect's control over our life. For a seventeenth-century brain scientist, Descartes was already way ahead of his time. Or, to put the same point the other way round, when it comes to understanding the human mind and human action, present-day brain science—despite its impressive accomplishments—still remains fundamentally Cartesian.

Even quite early in the history of modern philosophy, materialists (such as Hobbes, Lamettrie, and Diderot) tried to preserve the Cartesian view of matter without the Cartesian view of mind. But they too rejected only Descartes' substance dualism, while essentially retaining the Cartesian conception of what mind is like. It is that deeper struggle against Cartesianism that is most needed, but this is also where the greatest confusion reigns. John Searle and Daniel Dennett, for instance, can find no more devastating attacks on each other's philosophy of mind than by accusing each other of latent Cartesianism (Searle, 1992, p. 149; Dennett, 1993). It is easier, moreover, to find both accusations convincing than to decide how far the reluctant Cartesianism of each philosopher is responsible for what is false in each view and how far it is responsible for what is true. To me, it looks as if the animus against Descartes is due chiefly to the fact that he discovered (or invented) the modern problems of subjectivity and the mindbody (or even the mind–brain) relation, which we are still so far from solving that philosophers and scientists on all sides naturally project on him the discontent they feel toward their own frustrated efforts. Just as the first thing a recovering alcoholic must do is learn to respect the power that demon rum has over them, so the first thing we recovering Cartesians need to learn is respect for Descartes as the founder of all modern philosophies of mind.

The conception of mind that makes us all reluctant Cartesians is not Cartesian substance dualism, but rather a certain picture of the mind and its relation to the world. This picture makes of the Cartesian I a special kind of place—a metaphorically "inner" place, in which the I reposes at the center of its own little inner world of ideas. Outside this metaphorical inwardness—but now, strangely, the "outside"

seems literal, rather than metaphorical—there are objects, material bodies, including the body that gives the metaphorical inner place of the I its literal spatial location (for Cartesian materialists, inside the head, in the brain). Or rather, there are such external objects if, somehow, we can convince ourselves on the basis of these inner contents alone that there is anything outer at all—a problem which, in our overconfident spells of sobriety seems silly, but at other times formidable and even insoluble, when we see it through the bottom of our Cartesian bottle. Other I's, if there are any, have to be conceived of first and fundamentally as a species of these objects, even though they could not possibly be anything like them. Rather, they too must exist as other metaphorically inner places, whose existence we have to infer based on constructions or causal inferences from the behavior of certain of these (already always doubtful) material objects. Such a picture makes the mentality of others doubly unavailable to us: first, because it can never be available properly, in the way our own mentality is; and second, because it necessarily fails to partake of the only kind of objectivity—that of material things—that the mind is capable of grasping in contrast to its own subjectivity.

One take on the basic problem with the Cartesian picture is to say that it confuses mentality with perspectivity. For Descartes, a thought is something that is in us in such a way that we are immediately and infallibly aware of its presence, and also of its nature, at least when our attention is directed toward it (Descartes, 1964–1976, 8:7). This in effect supposes that the true nature of a mental state is grasped only from the “inside”—that is, as if the only perspective on it that counts is that of the person whose mental state it is. Mental states become inherently “subjective” in the sense that their true nature can be known only from a single point of view, that of the *subject* of these states.

It is as if we thought a smile or a grimace could be correctly apprehended and interpreted only when felt by the person who is smiling or grimacing, and that others must be dependent on the smiler's or grimacer's perception and interpretation for any correct information they might have about these facial expressions. The back of my coffee cup, that is visible to you and not to me, is as much a part of it as the front side, which is visible to me and not to you. Why not acknowledge that it is the same with mental states? For in fact my annoyance or rage are just as much present to you as the grimace or the angry outburst through which they are manifested. You understand the grimace only by understanding the annoyance, and vice-versa. To think otherwise is to misunderstand not only the mentality, but also the behavior. For there is no such thing as “first-person authority” regarding the content of our expressive behavior. The way I interpret the thoughts you express is as much a part of the nature of those thoughts as the way you intend them in expressing them, and the way you perceive my anger or sympathy is as much a

part of their nature as the way I perceive them. And this is even more true of our subtler and more complex thoughts than it is of our outbursts of joy or anger, for the reality of our thoughts consists as much in how others correctly interpret our expressed thoughts as in how we interpret them from our (perhaps interested and self-deceiving) perspective.

Descartes regarded thoughts as entirely real, in fact, as the only realities on the basis of which we could come to know about any other realities at all. There is nothing wrong with adopting the perspective of the subject for epistemological or methodological purposes, as Descartes did, and as Fichte did too, in making the active I the first principle of transcendental philosophy. But we immediately go astray about the nature of mind if we think that the nature of mental states can be grasped only from one perspective. For it belongs to the very concept of what is real or objective that it can be known from different perspectives, and its nature cannot depend solely on what is known about it only from one perspective. So the Cartesian conception of mind easily leads us to question whether mental states are part of the objective world at all. Do we need to study them in our psychology, or even mention them in a scientific theory of the objective world?

This last thought, as adapted by empiricist materialism, eventually led to what one can only describe as an utterly demented thought, which was nevertheless accorded honorific scientific status by a certain influential school of psychology (though it is less influential now than it used to be). This is the lunatic thought that psychological science must ignore the mental entirely and instead treat solely the externally observable physiological processes and bodily behavior of human organisms. The appeal of behaviorism and its more recent neuroscientific cousins to otherwise sane and highly intelligent people of scientific temperament is surely a devastating *reductio ad absurdum* of the Cartesian conception of mind that obviously lies behind such views.

Yet is it hardly less insane for me to think that the cordial or angry feelings you are showing towards me, or the thoughts about history or politics or philosophy you are communicating to me, really consist in neural processes going on inside your skull? For that implies that if I find your thoughts ambiguous or in need of clarification, or if I am unsure about the genuineness of your emotions, then all my questions on this score would best be answerable, at least in principle, by performing brain surgery on you and finding out what is literally going on inside your head. Someone who thinks that surely belongs in a padded cell just down the hall from those patients who think mental states aren't part of the objective world at all.

The real problem—not an unfamiliar kind of problem in philosophy, after all—is that, led astray by a certain picture, we can't understand any longer the very concept of the mental—at least not at the theoretical or scientific level—and so we

don't know where the concept of the mental states of others fits into the repertoire of concepts in terms of which we think, and must think, about things in general.

Because the noises I have just been making—about being trapped in bottles or pictures, and so forth—no doubt sound suspiciously like those customarily made by the later Wittgenstein and his followers, I must hasten to add that I regard the Wittgensteinian approach to mind as far worse, worse by something like an order of magnitude, than the Cartesian, materialist, and behaviorist ones I have been discussing so far. For the Wittgensteinian apparently thinks that everything would be fine and dandy just as it is, if only we could find some way to cure ourselves of our misguided tendency to ask philosophical questions about the mind. That thought, however, is reprehensibly complacent and obscurantist. For, of course, we need a correct theoretical understanding of the mental, and even a science of it, if we can get one. And such a science has to be integrated into our understanding of matter, biology, neuroscience, and the behavior of organisms. The problem is that all our attempts at a science of the mind up to now seem to be incurably Cartesian in ways we all know is deeply wrong but still have no idea at all how to correct.

8.3 A Transcendental Approach to Intersubjectivity

Well, perhaps it is a bit of an overstatement to say “no idea at all.” Or so I will try to argue. For I think that at least the first beginnings of a better theory of mind were made in the German idealist tradition, and some of them have been advanced both by Hegelianism and phenomenology. The origin of this tradition, as I have said, I believe to be in the thought of the founder of post-Kantian German idealism, Johann Gottlieb Fichte. We are still far enough from success in the attempt to de-Cartesianize our concept of the mental that Fichte's approach continues to have something to teach us.

What I want to discuss here is not so much a post-Cartesian conception of subjectivity or mentality in general, as rather a certain crucial aspect of it—namely, intersubjectivity, or our conception of the mentality of others and our awareness of it. One important approach to this topic developed on the European continent has been the phenomenological one, deriving from Husserl, and such followers as Edith Stein and Emmanuel Levinas. One possible weakness in the phenomenological approach is its way of answering the fundamental transcendental question, which Kant called the *quaestio quid iuris*, regarding the mentality of others, namely: with what right do we employ the concept of a mental state that is not our own? Some of Husserl's remarks might suggest that phenomenology's answer to this question is a kind of verificationist answer. By this I mean that it justifies the

employment of this concept by means of the criteria for the fulfillment of intentions directed toward the mental life of others (Husserl, 1960).¹ To the extent that this is Husserl's procedure, he recommends delineating the relevant experiences of intentionality and fulfillment, and once we see that the essentially pre-delineated conditions have been satisfied, and he thinks there is no further basis for any skeptical questioning about whether our experience really contains the objects of the specified intentions. Any verificationism of this kind would make phenomenology curiously uncritical, since it would seem possible to validate through such a phenomenological strategy any number of dubious concepts to which very real *quaestiones quid iuris* might rightly be raised. For people who believe in such things as streaks of luck, witches, or demonic possession, also become conscious of these supposed objects through a determinate species of intentionality, and their circle of illusion and self-deception surely also involves experiences of fulfilled or frustrated intentions that could be employed to construct verification procedures for what is objectively only a bunch of superstitious hokum.

The problem, from the standpoint of a more Kantian species of transcendental philosophy, is that such a verificationist strategy can never answer the doubts that might be raised *de jure* about the applicability of a concept whose role in experience is facing a skeptical challenge. Skeptics know that we are accustomed to employ such notions as material body, causal power, and the thoughts and feelings of others; they know as well as we do what we normally count as criteria for applying these concepts. Their claim is that we have no way of excluding the possibility that these procedures, and indeed any employment of these concepts at all, are simply based on an illusion. In this connection, Fichte imagines someone who replies to solipsism (he calls it "egoism") by appealing to everyday empirical criteria; he dismisses such a reply as "superficial and unsatisfying, . . . not an answer to our question at all Egoists also have these experiences to which appeal is being made" (VGB 6:303).

A properly transcendental answer to the *quaestio quid iuris* must follow after the manner of a Kantian transcendental deduction. Such a strategy sets the bar much higher than any verificationism, and answers skeptical worries the phenomenologist leaves unaddressed. A transcendental approach begins, namely, with some conception of possible experience that is so basic that it is very difficult to subject it to skeptical doubt without undermining the interest of the doubts themselves.

¹ See Fifth Meditation, especially § 52, in which the "style of verification" appropriate to experiencing the mental states of others is discussed. However, as I will mention presently, there are signs elsewhere in Husserl of an approach to the question that is much closer to Fichte's and not open to the objections I am presenting here. I am grateful to Dagfinn Føllesdal for bringing this more Fichtean side of Husserl's thinking to my attention.

Starting with a minimal and fundamental conception of experience, the transcendental philosopher argues for our entitlement to employ other concepts by showing them, and even their instantiation within experience, as a condition for the possibility of even this minimal starting point.

There are signs in some of Husserl's *Nachlass*, however, that he is actually much closer to this approach, where he argues that solipsism is self-undermining. This is certainly the way Fichte conceives of the transcendental problem of other I's. "We ourselves," he says, "introduce [other rational beings] into experience. It is *we* who explain certain experiences by appealing to the existence of rational beings outside of ourselves. But *with what right* do we offer this explanation? The justification needs to be better demonstrated before we can use this explanation, for its validity... cannot be based simply on the fact that we make use of such explanations" (VGB 6:303). Fichte proposes to deduce transcendently from a minimal standpoint—that of the self-positing I—not only the material world, and the necessary material embodiment of the I itself, but also the presence in experience of other I's, all as fundamental conditions for the possibility of the I's own original activity. I cannot, Fichte says, bring into existence I's other than my own. Yet "the concept of such beings underlies [my] observation of the not-I, and I expect to encounter something corresponding to this concept" (VGB 6:304). The point of transcendental argument is to justify this expectation, and to provide transcendental grounding for the concepts it involves.

Regarding the external world, Kant's Refutation of Idealism (in B) argues that the temporal determination of even the inner contents of the Cartesian mind would not be possible unless those contents were related to a material world distinct from them (KrV B274–9). Fichte presents a similar but distinct argument for the same conclusion, claiming that the "ideal series" of representations presupposes a "real series" of causal contacts between the I's material body, which is the indispensable vehicle for its activity, and a not-I or material world on which the I acts. Analogously, Fichte's strategy regarding other minds will be to claim that some of the essential activities of the I itself—in particular, those through which it determines itself as an individual agent—presuppose interaction with other I's and therefore the employment of concepts of an I that is not our own, as a condition for the possibility even of our own self-awareness.

Of course, a strategy of this kind is notoriously problematic, and arguments along these lines are famously difficult to execute. Philosophers outside the Kantian tradition have had no difficulty coming up with reasons for doubting not only particular transcendental arguments for such ambitious conclusions, but even the entire enterprise of transcendental philosophy itself. This might make it seem like a rather pointless exercise to provide such answers to skepticism, since

it is easy enough for anyone tempted by skeptical doubts in the first place to renew them by challenging the transcendental procedure. But this objection ignores the fact that for transcendental philosophy the real point was never merely to have an answer to skepticism, but rather to use this way of answering skepticism in order to provide insight into the nature of the fundamental concepts about which we are inquiring, and developing a new and revolutionary theory of the relationships between them. Skepticism can always persist, if it is stubborn enough. The value of skepticism to the non-skeptic is that it provides a challenge to our way of thinking about our fundamental concepts, and thus possibly an incitement to think about them differently. This is the way the transcendental philosopher really sees the matter. In the present case, I regard the real point of Fichte's transcendental approach to intersubjectivity as helping us develop an entirely new concept of mind, finding a way out of the Cartesian conception that we all find so unsatisfactory and yet also so difficult not to relapse into (at least on our bad days).

The fact that Fichte's first principle is the I has resulted in his being thought of by a long, ignorant, and uncomprehending tradition to think of him as just another Cartesian. The whole point of a transcendental strategy, however, is to begin with what your opponents can't plausibly deny, and then try to show that this commits the opponents to things they thought could be avoided. Thus a starting point that *looks* Cartesian is naturally used as a *way out* of Cartesianism. Fichte's I may be Cartesian in its certainty and its immediately and freely self-produced self-awareness. But we soon see that it is distinct from the Cartesian I in that it is limited to that direct awareness of awareness that consists solely in the free activity that is, according to Fichte, the essence of all subjectivity, and also what is most deeply fundamental in all experience. Describing his first principle, Fichte says: "Your thinking is an *acting*. Do not fear that through this admission you are conceding something that you will regret later. For I am talking only about the activity, of which you become conscious to yourself, and only to the extent that you become conscious of it" (ZE 1:522). The I "is only active, and absolute, never passive... it also has no being proper, no subsistence, for this is the result of an interaction... We should not even call it an active something, for this expression refers to something subsistent in which activity inheres" (EE 1:440). The Fichtean self-positing I is simply an activity—it is not even an active *something* (such as a Cartesian *res cogitans*), which, as Hume and Kant had already observed, is not given to us at all merely in the active consciousness of consciousness that constitutes the ground of all experience.

The next objection will naturally be that there is something incoherent and impossible about the Fichtean I, the thought of an activity without an agent—in other words, that the Fichtean principle of the self-positing I is necessarily an

incomplete thought, one that requires other thoughts, such as the thought of the *being* that is self-active. This point, however, should not be seen as an objection to Fichte, for it is precisely the point of a transcendental theory of experience to show that what we start with is possible only on the condition of what we transcendently deduce from it. Analogously, Kant proposes to show in the Transcendental Deduction that the synthesis of apprehension (a manifold of subjective representations in time present for a single, self-identical subject) is also not a complete thought, since it is not possible without syntheses of reproduction and recognition, which in turn are not possible without transcendental apperception and its relation to the concept of an object distinct from these representations and thought through the categories. And in the Analogies of Experience he proposes to show that a manifold determined in time is not possible without reference to the experience of a world of substances distinct from representations whose successive states are determined in time through causal laws and whose simultaneous states in space are subject to reciprocal causality. In other words, Kant's point is that the thought of a series of subjective states through time is an inherently incomplete thought, which requires for its completion a law-governed objective world. In the same way, it is precisely the point of Fichte's argument that the thought of the self-positing I is an incomplete thought that drives us on transcendently to think a whole series of necessary thoughts that are required for its completion. The only question is what these thoughts are and in what order they are necessitated.

8.4 The Other I as Condition for the I's Individuality

From the beginning of Fichte's attempt to develop a Doctrine of Science (*Wissenschaftslehre*) in 1794, it is clear that he regards the awareness of other I's as one of these necessary thoughts. In the *Grundlage* of 1794, he declares: "No thou, no I; No I, no thou" (GWL 1:189). But Fichte began to develop transcendental arguments for the other I only in the *Foundations of Natural Right* (1796), and he tried fully to integrate the theme of intersubjectivity into his Doctrine of Science only in the second main stage of thinking about his system, as it emerged late in his Jena period (1797–1799). This includes the first and second new *Introductions to the Doctrine of Science*, the *System of Ethics*, and the lectures customarily named *Wissenschaftslehre nova methodo*.

There the transcendental ordering of thoughts begins with the I's self-positing (or as he also calls it, "self-reverting" activity) and argues that forming a concept of this activity requires distinguishing it from an opposed activity, that of the object or "not-I" (ZE 1:492; cf. NR 3:17–28; SL 4:89–93). (The "not-I" spoken of in this

passage is as yet indeterminate as to what way it differs from the I, but Fichte's first exploration of the "not-I" considers it as the material world, or what differs fundamentally from any I. Later Fichte will argue that in order to be aware of my individual I, it is necessary to contrast it with another kind of "not-I," namely an I other than myself.) We first acquire our concept of the I as an acting being through the concept of its interaction with the not-I (the material world). The I as an acting *thing*, therefore, must also be material, a body (ZE 1:495; NR 3:56–61). The body is not an empirical accompaniment but a transcendently deduced requirement for being an active I at all: "Experience could not teach us that we have a body. That we have a body and that it is ours is something we have to know in advance, as a condition for the possibility of experience" (FGA 4:2:169). A disembodied Cartesian thinking substance is therefore incoherent, a transcendental impossibility. "Apart from connection with a body [an I] would not be a person, but would be something quite inconceivable (if one can still refer to a thing which is not even conceivable as 'something')" (VBG 6:295).

An active I "finds itself" (to use a favorite expression of Fichte in this connection), (NR 3:33; 4:18–21, FGA 4:2:181–2) only as *willing* (SL 4:18), and its willing takes the form of a striving against a material world on which it acts, and towards which it is at the same time also passive. At a general, indeterminate (and also unconscious) level, this willing is already present at every point at which the I finds itself, in the form of an indeterminate striving (GWL 1:65, 261–71, 285–91; FGA 4:2:77). But the I's awareness of itself as active is also an awareness of its activity as the determinate activity of this I. And it is at this still very fundamental point in the transcendental deduction of the conditions for the possibility of being an I that Fichte regards it as necessary to form the concept of other I's besides ones own, and to expect to encounter them in experience. For it is only through the experience of a certain kind of object, which is essentially distinguished from all *merely* material objects, that the self-consciousness of the I as a determinate form of activity can be thought of as possible at all.

The basis of Fichte's argument is the "self-limitation of the I," or the fact that any I is aware of itself not merely as rational activity or desire in general, but as an individual I freely choosing determinate modes of activity and having determinate ends or goals. At times, Fichte seems to mount a very short argument for his conclusion, inferring from the I's individuality directly to the necessity of opposing it to other individual I's:

Now the I is characterized by a free activity as such; hence the free activity must also be limited. That free activity is limited, means: one quantum of it is opposed to free activity in general, and to that extent to other free activity. In short, the I can appropriate to itself no free activity absolutely, without this activity being one quantum; and consequently,

without immediately positing simultaneously with that thought other free activity, which to that extent does not apply to it—since, indeed every quantum is necessarily bounded (SL 4:218–19).

Elsewhere (in a text that dates from about the same time), he uses similar reasoning to argue that Kant's unity of apperception cannot involve any individuality on the part of the I, which (he claims) can arise only when one I is thought of as opposed to others: "Nor can Kant understand by this pure apperception the consciousness of our individuality, or confuse the one with the other; for the consciousness of individuality is necessarily accompanied by another consciousness, that of a 'thou,' and is possible only on this condition" (SW 1:476).

8.5 The Summons

Fichte himself realizes that this argument taken by itself is too short to be convincing: "From this alone, however, the positing of individuality would not follow, for it would indeed be possible that the I should posit that free activity outside its own solely through *ideal* activity, as a merely *possible* one—possible to it itself, if it contained something perhaps beneficial, or also to other free beings" (SL 4:219). I can sufficiently individuate my individual I, in other words, without supposing actual rational beings outside myself, by thinking of alternative *possibilities* to my individual agency, adopted either by other possible beings or even by myself. But he does not find this reply decisive either, since it ignores the fact that what I must account for is not merely *thinking* of my I as an individual, but *finding* myself as an individual, as a determinate object. This finding, Fichte claims, is an incomplete thought (in the transcendental sense already mentioned), and is not possible without an *experience, as an object*, of the free activity constituting my individual I; and this experience is not possible without the instantiation in my consciousness of a distinctive concept, which Fichte calls the concept of a "summons" (*Aufforderung*). And this is a concept, he argues, that requires for its completion a reference to its origin in a rational being other than myself.

Here it is important to realize that what I have to be aware of in order to be aware of my individual I is not so much activity in which I have already engaged as *possible* activity in which I *might* engage. For me, my individuality consists not merely, and not fundamentally, in *facts* that distinguish me from others but in *possibilities of acting* through which I actively determine who I am. In other words, the awareness of my individuality must be fundamentally *normative*. This is so even if, as Fichte believes, the fundamental norms of rational agency are every bit as objective or universally binding as the objective truths about the material world. For

what makes me an individual is the specific way I apply these universal norms to my actions, in my unique situation and from my unique perspective.

Thus Fichte connects the concept of my individual I with the concept of a determinate *end or goal* (*Zweck*) of activity, and he regards the concept of an end as that through which an individual I can make a transition from (as he puts it) "determinability to determination" (SL 4:157; cf. FGA 4:2:47, 57–8, 175–6); that is, from the plural possibilities of what I might be to the free decision that determines who I am. This transition involves setting an end for myself, constituting my reason for making this transition, and then actualizing that end. The concept of an end, therefore, involves that of a possible way of acting that I have a rational ground for making actual. It is this concept, Fichte argues, that can present itself to me only in the form of a "summons"; and the concept of a summons is something that requires the actual existence of another rational being as its explanation. "The summons would thus contain within itself the real ground of a free decision; i.e. it would be the determining agency that intervenes between what is determinable and what is determinate" (FGA 4:2:179). "How is the concept of an end possible?" (FGA 4:2:173). "The end is given to us along with the summons . . . From this it follows that individual reason cannot account for itself on the basis of itself alone" (FGA 4:2:177). "I never find myself except insofar as I find myself summoned to act freely" (FGA 4:2:184). Therefore, "consciousness begins with consciousness of a summons" (FGA 4:2:189).

In other words, my own self-consciousness begins with my consciousness of another's consciousness as addressing me ("summoning" me). The sense in which Fichte means it "begins" with this consciousness is, of course, primarily a transcendental one. The concept of a summons is transcendently deduced as what is required for me to become conscious of my I in its individuality. But, of course, Fichte also clearly regards the "summons" as a kind of experience that we find in ordinary consciousness, paradigmatically, one supposes, when another person addresses you in a manner requiring some response. The recognition that a summons is necessary for individual self-consciousness means that the mental states of others, as perceived by someone other than the I whose states they are, are as transcendently necessary to the self-consciousness of an I as are its own states. Fichte's argument is that transcendently, the nature of mind is constituted as much by my awareness of the mentality of others as by my awareness of my own mentality.²

² From the beginning, Fichte writes as if the method of transcendental philosophy is what we could call a "linear" method, beginning with a first principle (the I) and progressing one step at a time, in a necessary order, from each transcendently deduced concept to the next. But we need to recognize that Fichte always regarded his system as something in which there would be a constant

In order to understand this argument, we need to understand more precisely what Fichte means by the word “summons” (*Aufforderung*). This term in German is broader in its meaning than the English “summons”—which perhaps suggests a command of some kind, such as a legal subpoena to appear in court. The term *Rechtsaufforderung* does cover that notion, but *Aufforderung* means something much less official and usually not in the least tinged with coercion. In many contexts, it is better to translate it as “invitation,” as in Carl Maria von Weber’s *Aufforderung zum Tanz*, op. 65. German dictionaries report that the meaning of *auffordern* ranges all the way from *bitten* (request or beg) to *verlangen* (demand or require). Fichte’s use here seems rather closer to the former end of the spectrum than to the latter. For he specifically emphasizes that an *Aufforderung* (as he means it) leaves us free *either to do or not do* as we are summoned (or invited) to do. “Either I act in accordance with the summons or I do not act in accordance with it. If I have understood this summons, I can, of course, still decide not to act [as the summons represents]” (FGA 4:2:179; cf. NR 3:34).

The whole point of the summons, in fact, is that it is what first makes our individuality possible for us, through presenting us with the concept of our own individual free action in the form of an object of our consciousness. I think Fichte chooses the term *Aufforderung* because its meaning is delicately balanced between the idea of something we merely can do and the idea of something we *should* do, or at least have some reason to do. What is clear, however, is that it *cannot* mean something we are *compelled* to do or have no choice about doing. Thus if the translation “summons” suggests legal coercion, then that is positively misleading as to Fichte’s meaning.

The decisive difference here is between an object that merely restricts our freedom and an object that makes freedom possible. This for Fichte is what is most basic to distinguishing the concept of another I from the concept of the mere not-I (the material world). The not-I (as the material world) resists our ends or may be brought into conformity with them. It may compel us to take one means to them rather than another, or it may make them impossible. But it cannot be the source

“back-and-forth” checking procedure, in which even the starting point would need confirmation by the way in which it makes possible a complete and satisfactory theory of experience (GWL 1:54, 58, 61–2). This implies, I think, that each stage of the transcendental procedure is also necessary in some way for the earlier stages—so that the summons is as necessary for our awareness of the not-I in general as the awareness of the material world is for our awareness of ourselves as embodied and for our awareness of the summons as addressed to us by another embodied I. It may also be relevant here that Fichte never actually brought any version of his system to the completion he projected, so that we have no actual instance in his writings of the kind of systematic structure he was seeking. The issues I have just been discussing, of course, raise questions about Fichte’s method that go far beyond the scope of the present topic.

from which we draw the concept of those ends. A summons, however, is precisely an object of consciousness that makes the concept of an end possible. It is, if you like, another species of not-I, a kind of not-I that does not resist our freedom, but on the contrary, is what makes freedom possible.

How can a summons, in this sense, be considered a transcendental condition of free activity? To act freely is to act in response to *reasons*. Reasons have the peculiarity that they are the only possible *determinant* of what we do that does not compel or causally necessitate what we do, or restrict in any way the possibilities we have open to us. A good reason explains why I do what I have reason to do, but never takes away from me the possibility of doing otherwise. In fact, it makes sense *as a reason* only as long as this possibility exists. Accordingly, there are two fundamentally different ways that facts in the world might be given to us as agents: first, there are facts that causally necessitate what we do, restricting our freedom to do otherwise; second, there are facts that determine what we do by presenting themselves as reasons for acting. I think Fichte was struck by the fundamental importance of this difference, and inferred from it that there must be something quite distinctive about the way that facts are given to us as reasons. His bold thought is that such facts can be given to us only through a distinctive kind of not-I that we regard as containing within itself the understanding of a reason, and hence free activity—in other words, through a not-I that is itself an I, namely, an I other than my own I. Only another rational being would be capable of having the concept of a free action and a ground or reason for free action. This is in fact the claim through which Fichte establishes this part of his argument.

I could therefore find a certain self-determination only through ideal activity; through imitation of one that is present at hand, and present at hand without my doing (*Zuthun*) . . . I cannot comprehend this summons to self-activity without ascribing it to an actual being outside me that wills to communicate a concept of the action demanded, and hence is capable of the concept of that concept; but such a being is a rational being, one that posits itself as an I, hence an I (SL 4:220–1).

At times, Fichte gives this last point what we may call a *genetic* presentation: being an individual I, placing before oneself an end, is something a rational being must be *educated* to do, through the influence of another rational being. “A human being becomes a human being only among human beings.” Freedom is possible only through *upbringing* (*Erziehung*) through the influence of other free beings (NR 3:39–40; cf. SL 4:221).

The summons should be understood as that kind of object through which something like a reason for a free action can first be given to us. Fichte's argument is that application of the concept of another I is the transcendental condition for the possibility of our awareness of a reason for acting.

It follows that if there are to be human beings at all, there must be more than one... The concept of a human being is not the concept of an individual—for an individual human being is unthinkable—but rather the concept of a species (NR 3:39).

Self-consciousness therefore originates with my act of selection from a general mass of rational beings as such... [A free individual] subsists only in the whole, and by means of the whole, as a portion of the whole (FGA 4:2:177).

Acting rationally, even acting autonomously, in other words, is not something a human being could do alone. Autonomy thus consists not in rejecting the influence of others, but in being influenced by others in the right way. Education, and being given reasons for action, constitute an essentially different way of being influenced by the world from any merely causal influence, through which one may be coerced, or manipulated, or conditioned to behave, but not enabled to act freely or autonomously. If we embrace some conception of mind and action that cannot distinguish what Fichte calls a “summons” from being causally influenced in general, then we should not expect to understand human freedom or rational action at all. Fichte’s argument implies that those who think of human individuality and freedom as somehow distinct from, or even in opposition to, human community, understand neither the nature of individual freedom nor the nature of community.

Fichte’s view here, if correct, would also have some important implications for our conceptions of reason and rationality. Giving oneself a reason for acting is derivative from being given a reason by others and from giving others a reason. Giving others a reason is the internalization of being given a reason by another, and giving oneself a reason is only an application to oneself of giving others a reason. Just as the nature of a mental state is not known exclusively by its owner, so a reason for me is not something answerable only to my perspective. Kant is right that rational thinking is thinking for yourself, but also from the standpoint of everyone else (KU 5:294–5; Anth 7:200, 228–9; VL Ak 9:57). It follows that there has to be something fundamentally wrongheaded about developing conceptions of rationality that are oriented exclusively to the agent’s standpoint (the agent’s desires, beliefs, and preferences). The ideally rational person cannot therefore be conceived of (as often seems to happen in the theories of rational choice theorists, game theorists, and economists) as a calculating sociopath with a gambling addiction (or, what amounts to the same thing, as a successful Wall Street investment banker in the underregulated financial sector of the U.S. economy).

Fichte’s approach to intersubjectivity also has important implications for the theory of human communication, and for the nature of language. Fichte developed some of these himself, both in his *Lectures on the Scholar’s Vocation* (1794) and in a brief essay on the origin of language written the following year. Our drive for rationality is transcendently connected to our drive to find other rational

beings in our experience, and to communicate with them, with the aim of achieving common agreement on the truth. This involves reciprocal effect or "interaction" (*Wechselwirkung*), which means both influencing them to accept our own views and being open to their influence (VBG 6:308). The same drive, Fichte argues, gives rise to the very possibility of sharing thoughts with others through meanings understood in common. Fichte thus goes beyond the Cartesian and Lockean semantic theory based on the premise that linguistic signs originally refer to "ideas" perceived by an individual mind, replacing that theory with one whose foundation is the transcendental claim that meanings are necessarily created by the I in its interaction with others. The idea "of indicating [our] thoughts through arbitrary signs, in a word: the idea of language. Hence the drive to find signs of rationality outside themselves harbors the particular drive to create a language" (FGA III/1: 103). As with many of Fichte's insights, this one was left to others to develop. It was the inspiration for the Romantic theory of language, especially as developed in the writings of August Wilhelm Schlegel, Friedrich Hardenberg (Novalis) and especially Wilhelm von Humboldt. According to this theory, the transcendental conditions for being an I are conditions for the possibility of an understanding between human beings that makes possible not only the means of communication, but even the desire to communicate. As Schlegel put it: "Not even the desire to communicate could be communicated if, before any agreed upon understanding takes place, humans did not already understand each other." (This tradition, and its origin with Fichte, is discussed by Mueller-Vollmer, 1989, pp. 162–84).

8.6 "No Thou, No I"

If there is to be an I as an individual mind, in an individual body, then there must be a plurality of I's, capable of addressing one another and rationally communicating with one another. We cannot comprehend the mind, or thinking, solely from the standpoint of the Cartesian meditator, sitting cozy and warm in his *poêle*, safe from the cold Dutch winter, but also isolated from the thoughts of others. The meditator's *poêle* is only the Cartesian metaphor for the mistaken Cartesian conception of the mind as an isolated "inner" place, and this is equally true whether we think of this place as an extensionless, windowless spiritual substance, a Humean theater where impressions and ideas disport themselves, or a site of a wondrous wealth of neural processes taking place inside the human head.

A human mind may not be possible without a human brain, but it is also impossible if we consider nothing more than that. We should take seriously, therefore, those neurophilosophers who claim that if we consider only neural activity, we must eliminate all intentional states, all beliefs, desires, and directed thoughts

(see Churchland, 1981, 1988). We should understand their message, however, not in the sense that these intentional states and activities do not exist, but rather in the sense that they cannot be identified merely with, or properly understood merely in terms of, neural processes.

To think and act, a human being must be brought up or educated to think and act, by human beings already capable of thinking and acting. Thought and action, therefore, can never be understood in their nature unless they are understood in the context of a community of rational beings, acting and interacting, understanding and interpreting one another, and offering to others communicative action, which these others can interpret. In short, “No thou, no I” (GWL 1:189).

We can also look at Fichte's conception of the intersubjective I as an insight about what some philosophers like to call the “first-, second-, and third-person standpoints.” The thesis is that the first-person standpoint is already implicitly (as a matter of transcendental conditions for its possibility) the second-person standpoint. To consider things from my perspective is already to presuppose other perspectives on mine, and my perspective on them: I can be aware of myself only by being aware implicitly of the possibility of other perspectives on me, and of my perspective on them. And every third-person (or supposedly “objective”) perspective, if it is a perspective on a human being, is also implicitly “subjective”—its object must be someone who could address you (who takes the “third person perspective”) and someone you could address in the second person.

This means there is nothing merely subjective or apparent about any of these perspectives. The conditions of first-person agency, for example, are no more fundamental (as Cartesian methodology would have it) than the conditions of second-person address, or third-person commentary. Fichte's own beginning with the I (with its apparent Cartesian privileging of the first person) is designed from the start to undermine itself: There can be no awareness of an I without awareness of a not-I (an objective world) and awareness of a “thou.” But equally, the first-person standpoint, with its awareness of practical engagement with the world and of its own absolute freedom, cannot be dismissed as a subjective illusion, merely “a way things merely seem,” like the apparent motion of the sun around the earth, or what some psychological literature shows us about our tendency to distort past experiences in recalling them (“cryptomnesia,” selective memory), misdescribe our inferential processes, or misjudge probabilities (e.g. the “gambler's fallacy”) (see Schick and Vaughn, 2002, Chapters 3, 6). All these “ways it seems to us” can be replaced by a more “objective,” third-person account of our mentality and experience—and they *should* be, even from our own first-person perspective, if we are to be rational subjects. But the same is not true of the first-person perspective itself as regards what it fundamentally presupposes if we are to make

coherent judgments at all (see Chapter 7, § 7). This can no more be dismissed as an illusion than that there are material things outside us, or other I's that summon us.

All three "perspectives," then, are related reciprocally, none is absolutely privileged, and all three are conditions of the possibility of any experience of an objective world. If there are tensions between them, these cannot be solved by dismissing any of them, but only by reconciling them. Or if, as some believe (see Nagel, 1989) this proves to be impossible, then we have to acknowledge that there is a permanent, insoluble philosophical problem, a fundamental existential absurdity lying right at the ground of our human existence. (Our consolation would then be that this is not the only such insoluble problem or absurdity, as every honest philosopher knows, a truth that only shallower minds—either scientific or religious—would attempt to deny.)

These are some of the implications of Fichte's transcendental deduction of the other through the concept of a summons. The main question, however, may still not seem to be sufficiently answered: why must the concept of an end, or of our possible free activity, be given to us from outside, through a summons? Why isn't it simply part of our *internal* equipment as free rational agents? But let us recall, first of all, where we began and how far (or rather, how short a way) we have come from that beginning. We began with the I not as a full blown empirical self, but simply as freely self-positing activity. This drove us to recognize certain other presuppositions or transcendental necessities required by this starting point. An I must be opposed to a not-I, or a material world, and it must therefore itself take the form of a material body. It must find itself as striving in opposition to this world. And then it must think of this striving as individualized through the concept of determinate possible activity and the corresponding normative concept of an end. The determinacy of self-positing in general requires the concept of something *external* to the I which is given in opposition to it—the not-I, the material world. Likewise, it seems reasonable to Fichte that this new determinacy, through which the I is individualized, should also be thought of as something given from outside.

Fichte's first thought (the "very short argument") was that the thought of one individual I requires that of other individual I's characterized by different free activities from that of this individual I. This argument, he says, can be avoided by suggesting that the different free activities might be thought of as only *possible*, in contrast to the free activity of this individual I. But, he claims, that raises the new question of how it is possible for us to form the concept of alternative *possibilities* of free activity as something "found" in experience, since these are also not yet contained in the concept of a self-positing I, embodied and striving against a material world. Fichte's claim is that the concept of such possible free activity amounts to the concept of a summons, a distinctive kind of object, in which the possibility

of free activity, and the givenness in the world of possible reasons for acting in that possible way, is present to us. That, in turn requires positing, as its source, other rational beings or I's distinct from and in relation to the individual I, since only such a being could contain in itself the concept of a possible mode of free activity or the concept of a reason for choosing it. The necessity and power of this argument, if I may report my own reaction to it, grows on you the more you think your way into the transcendental questions it is meant to answer, and the more you see how these questions undermine the set of Cartesian and empiricist prejudices about the nature of mind and free action that we find so hard to resist, at least on our bad days.

Of course to many, the whole project of transcendental philosophy in general, and therefore also Fichte's system of positings, counter-positings, and the system of concepts arising from it in particular, may seem like an arbitrary invention of fancy, rather than a necessary rational progression. If that's how it strikes you, then none of this argumentation for the necessity of the summons and of other I's will seem the least bit convincing or illuminating. But it is one of the chief purposes of transcendental philosophy, as Fichte does it, to change our way of looking at ourselves and our experience, to ask fundamental questions about them that we never thought to ask before, and in response to provide a more fundamental grounding to familiar concepts than we had before, and in the process to reshape and reorder those concepts. If Fichte is right, then the concept of other rational beings and our relationship to them plays a much more fundamental role in both our experience and our agency than was recognized by traditional philosophy, especially by the philosophy of mind we have picked up from Descartes and the empiricists—much as we might have picked up our alcohol addiction from our parents' bad habits and degenerate way of life. Fichte's transcendental arguments certainly provide us with no Twelve-Step Program, and any pretense that we can rely on a "higher power" is as fraudulent in the one case as in the other. Whether or not they are capable of putting us on the road to a better way of doing the philosophy of mind, at the very least they open up for us a new way of looking at things that might help us to a future of more good days than bad days.

9

Fichtean Themes in Hegel's Dialectic of Recognition

Hegel's theory of recognition in Chapter IV of the *Phenomenology of Spirit*, culminating in the famous master–servant dialectic, is one of the most famous of Hegel's contributions to philosophy. Its influence can be seen in such widely divergent philosophical theories as Marx's theory of alienation, Nietzsche's genealogy of morals, and Sartre's phenomenology of our experience of the other, and it still occupies philosophers today, for instance, in the recent work of Axel Honneth (Honneth, 1995, 2010). Hegel's dialectic of recognition contains some crucial ideas that are startlingly original, but it would not have been possible except for the even more original introduction of the very concept of recognition (*Anerkennung*) into philosophy by Hegel's great predecessor, Johann Gottlieb Fichte. My aim here will be to review Hegel's discussion in the *Phenomenology* that leads up to his dialectic of recognition, and identify the Fichtean themes which frame that discussion. My larger purpose is to enable us to appreciate Hegel's original contributions by setting them off against the Fichtean background that made them possible. This expresses my general approach to Kant and his successors, which is not to harp on the relatively minor quarrels and quibbles among them, but rather to emphasize the continuity between them that separates them from pre-critical metaphysicians, dogmatists, empiricists, and all others who have yet to rise to the point of even understanding the project they all share, or appreciating why it is indispensable to philosophy.

9.1 Transcendental Philosophy in Fichte and Hegel

Fichte and Hegel are, namely, *transcendental* philosophers. Their emphasis on subjectivity, as Frederick Beiser has emphasized in a recent book, is the most extreme

repudiation possible of the subjectivism that reigns in Cartesian and empiricist metaphysics (Beiser, 2002). Its point is never to deny the reality of the objective world, but rather to confirm it by deriving the necessity of objectivity from the fundamental character of what is necessary for us to have any consciousness at all. For both Fichte and Hegel, what is most fundamental is active self-consciousness—what Fichte most often calls the “self-positing I” and Hegel calls “the concept.”

For Fichte, during the phase of his Jena period when he best developed his theory of intersubjectivity and recognition, the starting point of transcendental philosophy consists in summoning his readers or hearers first to think an object—such as the wall in front of them, the desk they are sitting at, or the stove in the corner—and then to form a *concept* of that which has done that thinking (ZE 1:458, 525–31; FGA 4:2:28). He calls the result of this summons a *postulate*, I think in a fairly strictly Euclidean sense. Euclid's postulates are requests to perform an action (drawing a line between two points, transposing or translating an angle or figure, extending lines that are at right angles to a given line) and then to concede a certain proposition about the result of this action (that a single unique straight line may be drawn between any two points, that all right angles are equal, however translated or transposed, that parallel lines never meet). Fichte asks us to think of the I that thinks an object, and then to concede certain things about it. The first of these things is that the awareness of the I, like the awareness of the wall, the desk, or the stove, involves an intuition—a direct cognitive contact with what it is an awareness of. We are also to concede that the intuition of the I, unlike that of the object, is *intellectual*, rather than *sensible*; that is, it is due to a free act of ours, rather than to our being passive to something, and moreover it is an intuition solely of our *acting* itself, not of any *being* apart from or underlying the action (as is the sensible intuition of the wall or stove). This is due to the fact that the act through which we become conscious of the self, in contrast to the act through which we become conscious of the wall or the stove, has the distinctive character that it is “self-reverting” (*in sich zurückgehendes*): it is an act directed only at itself, its own acting, and not at any being distinct from that acting.

To form a concept of the I is, as Fichte says, to *construct* that concept (ZE 1:459; FGA 4:2:34), and he means this in the sense in which Kant thinks we construct the concepts of mathematical objects: we first exhibit that of which we form the concept a priori in pure intuition, and then we determine in thought that which we have intuited. Fichte argues that we cannot do this without simultaneously positing in opposition to the I some object or being or not-I (such as the wall or the stove) which is different from the I and which we intuit sensibly rather than intellectually. In this way, the existence of, and our awareness of, an objective world is shown to be a transcendental condition for the possibility of self-consciousness.

Thus every act of awareness—of the wall, for instance—is also twofold: it involves a sensibly intuitive awareness of the wall, and also an intellectually intuitive awareness of our own awareness, of our own acting in being aware. This means that all experience involves a twofold series, which Fichte calls, respectively, the “real” series—of awarenesses of the wall, the stove, my hand in front of me, and so forth—and the “ideal” series, of my acts in representing the wall, the stove, etc. These intuitions, of course, and especially the intellectual intuition of my own acting, need not be reflected upon or brought to conception separately. But when they are, they constitute a distinctive awareness, that of a *representation* (in the sense that had been brought to the attention of philosophers by Karl Leonhard Reinhold, when he proposed to ground all experience on “representation” as that which distinguishes itself from the subject and the object while relating itself to the subject and the object). Thus we see that both the objective world and the series of our subjective representations of it are transcendental conditions of self-consciousness. We also see that, in contrast to the Cartesian and empiricist metaphysics, the relation between these two series is not a causal one. We do not reach the real series by means of a causal *inference* from the ideal series: the truth, in fact, is closer to the exact reverse of this; the real series is transcendently more basic as a condition of the possibility of the self-positing I, and the ideal series simply a way of reflecting on the I's subjective relation to its reciprocal interaction with the not-I. Finally, the difference between the two series gives us no ground at all for skeptical doubt about the objective world of which we are conscious in the real series. Rather, the point is to provide a *transcendental deduction*, as a condition of the possibility of self-consciousness, of both the concept of an objective world posited as distinct from the I and of the concept of a series of representations of that world belonging to my reflection on my awareness of that world. I mean this in the Kantian sense that we have an answer to the *quaestio quid iuris* about these concepts: we have an a priori justification for employing them and an assurance a priori that they have instances in our experience.

When we turn to the opening chapters of Hegel's *Phenomenology of Spirit*, I think we find exactly the same aims, but the steps of the argument are carried out, in a sense, in just the reverse order. In the Introduction, Hegel does not begin with our act of forming a concept of the I, but rather with “natural consciousness,” which Hegel appears to equate with experience in Reinhold's sense, as *representation*: “Consciousness simultaneously distinguishes itself from something and at the same time relates itself to it” (PhG, ¶ 82). The being for consciousness of this thing, Hegel says, is called “knowing,” while the being in itself of the thing is called “truth.” (So we have already before us Fichte's ideal series and real series.) The procedure then is to execute a self-critical examination of this natural consciousness,

which repeatedly undermines its own concept of both the truth to be known and the way it regards itself as knowing that truth, and therefore successively transforms both these concepts.

Thus Chapter I of the *Phenomenology* grasps knowing simply as sensing, and the truth as sensuous immediacy. In Chapter II, knowing becomes “perceiving,” or attempting to equate a sensuously conditioned universal concept with an object determined by it, and truth becomes the self-sameness of the two. Then in Chapter III, knowing is transformed into forming concepts that are unconditioned by sense, and attempting to apply them to an object that is “supersensible,” not immediately available to sense experience but constitutes an “inner” reality that manifests itself through “appearance” regarded as the process of instantiating these concepts in some experienceable way. Chapter III then has a surprising outcome: that which has all along played the role of object or “other” to consciousness, when revealed in its truth, turns out to be nothing but consciousness itself. Or as Hegel puts it, when we draw away the curtain of appearance, what lies behind it is nothing but the self that has all along thought of the truth as precisely some object distinct from itself (PhG, ¶ 163). Thus the entire project and orientation of natural consciousness from the Introduction onwards is exposed as fundamentally false; truth is not to be found anywhere in it at all. “With self-consciousness, then, we have thus entered the realm where truth is at home” (PhG, ¶ 167). In this way, the starting point of Fichte’s transcendental deduction is for Hegel to be vindicated as the only possible outcome of a critical examination of the natural standpoint that seemed to constitute its total rejection.

This self-critical process, to which Hegel gives the name “experience,” has a twofold meaning: on the one hand, it displays the inadequacy of the shapes of consciousness through which experience passes; but on the other hand, it also provides a kind of transcendental deduction for the concepts involved in those shapes of consciousness, providing us (in proper Kantian fashion) with a deduction of their validity and also an exhibition of the proper extent and the necessary limits of their applicability. That is to say: immediate sensation, the process of comparing sense-universals with their objects, and the formation of theories framed in abstract universal terms about how reality manifests itself in experience—these are all ways of knowing truth, and each has its place. But they are not equal in providing access to truth. Immediate sensation is inferior to conceptualized perception, and this in turn is inferior to scientific theories guided by thought-determinations grasped in their a priori validity. Further, this whole orientation—the theoretical orientation, which identifies truth with an object and conceives of the subject of knowledge as a mere beholder of a world of objects—is exposed as limited and inadequate. On the contrary, truth is properly at home only

in the self-consciousness of the subject itself and, as we presently see, that means in its *active* relation to itself and to the objective world that theoretical knowing therefore grasps only in a limited and inadequate way.

This Hegelian conclusion, of course, is also Fichtean through and through. One of Fichte's chief aims in the *Foundations* of 1794 was to demonstrate that "reason cannot even be theoretical if it is not practical" (GWL 1:264). In the post-1797 version, he emphasizes that the fundamental principle of idealism is practical freedom, the intellectual intuition of acting is at the same time awareness of a practical capacity. Fichte's conception of the first principle itself as a *postulate* involves conceiving it as a self-conscious response to a summons to the free action through which the I constructs itself at the same time that it constructs its concept of itself (ZE 1:459–60).

9.2 Self-consciousness in Hegel's *Phenomenology*

Hegel's development of the standpoint of self-consciousness at the beginning of Chapter IV of the *Phenomenology* contains a number of otherwise mysterious moves that are best illuminated by seeing them as acceptances of a series of transcendental claims that Fichte makes about practical self-consciousness. Hegel claims that self-consciousness is originally "desire in general," whose content in general is that the unity of self-consciousness with itself should "become essential" (PhG, ¶ 167). Further, the coming to be of this unity, according to Hegel, occurs through a relation of self-consciousness to its external object. This relation, also called the "satisfaction of desire," is characterized as the "negation" of the object, the "superseding of its otherness" or the "destruction of its independence" (PhG, ¶ 174–5). Another, perhaps even more seemingly arbitrary, set of claims that Hegel makes about self-consciousness in this passage is its identification with "life," in a sense that implies a plurality of independently existing living things "existing for themselves" (PhG, ¶ 168–71). Yet life "as a process"—the process, namely, that unites these independently existing living beings, "points to something other than itself, namely to consciousness, for which life exists as this unity, or as a species (*Gattung*)" (PhG, ¶ 172).

In Chapter IV of the *Phenomenology*, these claims seem to come out of thin air. Only the most stubborn, fanatical, and mentally unbalanced kind of Hegelian would pretend to find in Hegel's text anything resembling *cogent arguments* to connect self-consciousness with desire, life, and the species process. But in Fichte there are some highly plausible lines of reasoning in support of Hegel's conclusions, or at least something analogous to them. For Fichte, the self-reverting (or "centripetal") activity that posits the I stands opposed to the objective (or "centrifugal") activity

that counterposits the not-I, as a necessary condition for the I's forming a *concept* of its own intellectual self-intuition. When these two necessary activities are considered in their practical significance, the not-I appears as an efficacy resisting the original activity of the I.

This leads, to begin with, to the inference that the I and the not-I must exercise efficacy on each other. This means that when the I forms a concept of itself not merely as an *acting* but as a *being that acts*, the same kind of causal efficacy must pertain to this being as pertains to the not-I with which it interacts. Thus the I must regard itself as embodied, and a Cartesian disembodied thinking subject becomes transcendently unthinkable (VBG 6:295; NR 3:56–61). Further, through arguments I won't attempt to summarize here, Fichte defends the position that this body must be regarded as an organic unity (an articulated living thing, an animal organism) (NR 3:63;85, SL 4:112–22). Fichte thus has a uniquely interesting solution to the problem posed by the fact that I must identify myself with my body, while also thinking of my own body as an object, hence distinct from the I. Since the I is not a thing or being but an activity, this makes it distinct from all beings, including even that very being with which it must identify itself when it forms the concept of itself as an active being.

9.3 Desire

Fichte also sees the causal interaction of the I and the not-I as posing another problem for transcendental philosophy whose solution is crucial to understanding the necessary conditions of practical reason. If we adopt the transcendental standpoint, we understand the not-I as first arising from the positing of the I; yet the not-I is what it is fundamentally because it opposes the activity of the I. How, then, can we conceive of the activity of the I that originally posits the not-I? Its relation to the not-I makes it a kind of causality, yet the causality of the not-I cannot collapse into the causality of the I without simply contradicting the concept of the not-I. This leads Fichte to characterize the activity through which the I posits the not-I in a paradoxical way: it must be “a causality that is not a causality”; that is, Fichte says, a *striving* (BWL 1:65, GWL 1:286). This striving is originally boundless, without a determinate object, and even fundamentally unconscious (GWL 1:265). We experience it empirically as a “longing” or “yearning” (*Sehnen*) (GWL 1:303–5), and Fichte locates this striving within the organic tendencies of the I's living body (SL 4:122–8). This Fichtean line of reasoning, if successful, would justify Hegel's association of self-consciousness with desire, and moreover with desire grounded in the life-processes of the self-conscious being as an animal being.

Hegel's conception of the satisfaction of desire is also based on Fichte. For Fichte, the not-I is present originally as that whose causality provides a "check" (*Anstoß*) on the activity of the I. All successful action, therefore, has the character of overcoming the not-I's resistance (and in that sense, "negating" the not-I). For Fichte, this takes the determinate form of conceiving for each object a corresponding "practical concept"—the concept of its "final end" (*Endzweck*) or of how that object ought to be in order to be brought into unity, harmony, or agreement with the I, and action takes the form of altering the object so as to subordinate it to this practical concept. In this way, the I's striving to be in unity with itself is mediated through its unity with the not-I (VBG 6:297–9; SL 4:170–2). The satisfaction of any desire, as regards its form, thus consists in subordinating an object to the I's concept of it, which Hegel represents as "negating" the object or "superseding its independence."

9.4 Intersubjectivity

Still more Fichtean arguments might explain Hegel's association of self-consciousness with a plurality, or even a "species," of independent living beings. Fichte thinks that contrasting the I with the not-I is necessary to form the concept of the I that thinks an object and acts in the world. But he does not think it is sufficient to enable me to form the determinate concept, which I must form, of my I as an individual subject of thinking, knowing, and acting. Kant, he claims, "cannot understand by pure apperception the consciousness of our individuality, or confuse the one with the other; for consciousness of individuality is necessarily accompanied by another consciousness, that of a thou, and is possible only on this condition" (ZE 1:476). The concept of a rational being, he says, is essentially the concept of a *community*, not of an individual (SL 4:64; VBG 6: 304–6, cf. NR 3:39). "If there are to be human beings at all, there must be more than one... Thus the concept of the human being is not the concept of an individual... but rather the concept of a species" (NR 3:39).

There are, I think, really two Fichtean lines of reasoning here. One is that I-hood, which Fichte identifies with rationality, is essentially a normative concept, and one whose norms are essentially universal in scope. To be an I is to cognize, act, and set ends whose common form is to unify—to make a coherent reality out of experiences, to give continuity to oneself through time through the coherent pursuit of ends and conduct according to universally valid principles. In this way, to be an I is implicitly to be only one of indefinitely many possible I's, all of whom are subject to the same normative principles. The Kantian unity of apperception, for Fichte, is merely the abstract thought of the kind of being that is subject to the laws of

understanding that constitute experience. Hegel expresses the same point when he notes that "I" is both the most universal and the most particular of concepts—everyone is an I, anyone at all is an I, yet "I" also refers precisely to the individuality of each subject (EL, § 20, R).

The second Fichtean argument takes us pretty directly to Fichte's theory of inter-subjective recognition, hence to the direct source of Hegel's theory of the struggle for recognition and the master–servant dialectic. This is the argument that positing oneself as an individual requires the setting of an end, and the original possibility of setting of an end is possible only in response to a "summons" (*Aufforderung*), which is experienced only as coming to the I from outside, from an object which is itself in possession of the concepts of an end and of free agency—in other words, from another I (NR 3:32–40; SL 4:218–22). There are, then, for Fichte, fundamentally two kinds of objects, two different concepts of the not-I. One is that of an object that checks or limits our agency, but the other is that of an object that does not oppose free action, but rather opens up for us the possibility of it. This is what makes the experience of a "summons" fundamentally different from the concept of any effect on us of a mere object, or even the effect on us of the actions of a rational being insofar as that effect is something that limits our free action, or something we either seek to resist and overcome or to utilize as a means for our other ends.

I suggest we formulate his thought as follows. The summons, requiring an attribution of it to a rational being outside me, is the transcendental condition of the possibility of my having anything like a *reason* for acting. In the following passage from his lectures, Fichte comes very close to saying exactly that: "The summons would thus contain within itself the real ground of a free decision; i.e. it would be the determining agency that intervenes between what is determinable and what is determinate" (FGA 4:2:179).

Fichte's claim is not, of course, that whenever I have a reason to do something, I must derive it from a summons by another person. It is rather that the capacity to give oneself reasons is something that needs to be acquired through interaction with others, just as the capacity to have theoretical knowledge of the world is dependent on the I's practical limitation by, and interaction with, the not-I. More specifically, giving oneself a reason for acting is derivative from being given a reason by others and from giving others a reason. Giving others a reason is the internalization of being given a reason by another, and giving oneself a reason is only an application to oneself of giving others a reason.

The capacity to give myself such reasons is not, in other words, part of the original equipment belonging to the intellectual intuition that constitutes I-hood, any more than is the capacity to distinguish the self-reverting (centripetal) activity of *self-awareness* from objectively directed (centrifugal) activity of *objective*

awareness. The latter requires us to posit (and thus provides a transcendental deduction of the concept of) a not-I, or an objective world in general outside me. The former requires us to posit (and thus provides an analogous transcendental deduction of the concept of) another rational being, with whom I stand in a relation of community, which Fichte characterizes as being able to give and receive communication about reasons to think and act that are answerable to universal normative standards (VBG 6:304–6; NR 4:163–76).

9.5 Subordination and Co-ordination

Just as the fundamental relation in which the I stands to the mere object is that of desire, the form of whose satisfaction is the *subordination* of objects to the I, so the fundamental relation in which the I stands to other rational beings is that of *co-ordination*, mutual rational communication and mutual education, leading to a relation of rational reciprocity (*vernünftige Wechselwirkung*) or community (*Gemeinschaft*).

The first condition for human community, according to Fichte, is the mutual recognition (*Anerkennung*) of rational beings as standing in a relation of right. Each rational being recognizes the other as free, and the fundamental summons contained in the relation of right is that each rational being has an external sphere of freedom, starting with its own body and extending to all those actions involving rightful use of its property, which other rational beings are to respect and not to violate (SW 3:41–7). This relation of recognition is the basis on which Fichte builds his theory of natural right. But in fact, Fichte's theory of intersubjectivity extends much farther, since he regards the state as enforcer of external right as only a necessary condition for the relation between rational beings that constitutes true human community (SW 6:306–7).

In this fuller relation, no rational being negates the freedom of the other, but on the contrary, each makes possible the liberation of the other, and each adopts its own ends within a systematic framework of *common* or shared ends, set and pursued in conformity with rational standards universally valid for all (SW 6:305–11). Of course, this is obviously a version of the Kantian idea of the realm of ends, but one that Fichte creatively develops into a marvelous and inspiring Enlightenment vision of the essential human vocation, which is the vocation of belonging to a free, rational community of equals. This vision is all the more inspiring because it shows how the Enlightenment ideals, the ideals also of the French Revolution—namely, freedom, equality, and community—are by their very nature not in conflict, but are only different aspects of the same ideal. Those who see ultimate tensions or dilemmas in pursuing them therefore reveal that they do not understand the nature of the freedom, equality, and community that they are pretending to pursue.

It is striking that during the later Jena period, when he develops the theme of the summons and mutual recognition, Fichte incorporates his theory of intersubjectivity even into the method of transcendental philosophy. For recall that the starting point of Fichte's *Wissenschaftslehre* at this stage is a postulate, a request, or invitation, or summons (*Aufforderung*): "Form the concept 'I' and observe how you do it." Fichte even calls attention to this feature of his method, by pointing out that his method can get started only through his addressing the reader in the second person (ZE 1:521); and several times he describes himself as "summoning" the reader to a free act of self-positing and self-construction, and of the reader who follows him as responding to a "summons" (ZE 1:445, 461–2, 531). Hence all those more recent philosophers who insist on repudiating "monological" models of thinking and incorporating communicative interaction with the other into philosophical method ought to acknowledge Fichte as the first philosopher to express their insight and the true originator of their entire approach.

9.6 Self-certainty and Recognition

This, then, is the Fichtean background of Hegel's theory of recognition developed in Chapter IV of the *Phenomenology of Spirit*. That theory is on the surface obviously very different from Fichte's, but I see here another instance of the same pattern I described earlier in comparing the two philosophers contrasting ways of establishing the active I as the transcendental foundation of theoretical cognition. Fichte's argument proceeds directly, from the I and the transcendental conditions for its self-conception as an active individual. Hegel's argument proceeds indirectly, from a standpoint that is apparently the exact opposite of the one on which Fichte builds his transcendental theory. But the ironic outcome of Hegel's argument is to confirm a conclusion very similar to Fichte's, yet in a way that is so strikingly original that it is all the more compelling.

Hegel's discussion begins with self-consciousness as desire, and the conception of satisfaction as negation of the object (PhG, ¶ 175). And instead of arguing, as Fichte does, that self-consciousness in its individuality requires a different kind of object, to which its appropriate response is co-ordination and communication, rather than subordination and negation, it regards consciousness as persisting as long as possible with the project of seeking satisfaction simply through negating the independence of the object. The concept of another self-consciousness as the object of desire is transcendently deduced instead from the inadequacy of any less complex kind of object to satisfy the desire of self-consciousness. Hegel's explicit argument here is that a simple object, once negated, ceases to offer any further satisfaction to self-consciousness. What is needed instead is an object that

persists in its being even after being negated. This, Hegel claims, can be found only in the form of an object that “itself effects the negation within itself,” and this in turn is possible only for something that is its own negative or its own object, namely, another self-consciousness (PhG, ¶ 175).

The argument here seems unconvincing, since it remains unclear why the round of “object → negation → satisfaction → new object → negation → satisfaction” should fail to satisfy the desire of self-consciousness. I suggest that Hegel’s argument makes more sense if we recall that the deeper aim of satisfying desire is to achieve what Hegel calls “self-certainty”—in other words, the vindication from outside of self-consciousness’s claim to be “essential” to itself (PhG, ¶ 174), or to have a certain kind of *value* for itself. Perhaps the only kind of object whose negation is capable of confirming or vindicating self-consciousness’s desire for self-validation is an object that is itself capable of explicit self-valuation, and that would have to be another self-consciousness. There thus arises for Hegel’s self-consciousness also an idea of community, of an “I that is we and a we that is I,” which Hegel describes as “the concept of spirit” (PhG, ¶ 177). But this community is one in which self-consciousness still regards the other only as an object of desire, as something to be negated in order to obtain its own self-certainty, the vindication of its self-valuation. This self-valuation, reflected back out of another self-consciousness, is what Hegel means by “recognition” (PhG, ¶ 178).

9.7 Struggle and Domination

The satisfaction of self-consciousness’s desire for self-certainty, then, consists in obtaining from another self-consciousness the negation of itself, which renders it dependent. Hegel now imagines two self-consciousnesses engaged in a struggle to the death, each with the aim of obtaining recognition without having to give it in return. By being willing to stake its life, each self-consciousness asserts the value of self-certainty to be greater to it than the value of life, and this amounts to a way of obtaining self-certainty. But the attempt to confirm this self-certainty through the death of the other necessarily fails, because the struggle can end only with the survival of a single self-consciousness, with no object self-consciousness left to recognize it.

Self-consciousness can therefore obtain self-certainty only in another self-consciousness that, through being overpowered by the fear of death, chooses life over recognition; in other words, a self-consciousness that is prepared to live by negating itself on behalf of another, to recognize another without being recognized in return, and thus to live as a dependent self-consciousness. This Hegel identifies as the self-consciousness of the servant (*Knecht*), while the self-consciousness that

obtains self-certainty through making this servant self-consciousness its dependent object is called the master (*Herr*) (PhG, ¶ 189). The relation between them shows itself in their respective relations to the world of simple objects of desire. The servant negates these objects by laboring on them, he turns them into objects of satisfaction, yet not of his own satisfaction but of the master's. The master obtains the satisfaction, but without performing the act of negation (PhG, ¶ 190).

There is an irony here, however, which Hegel hastens to point out. The master is supposed to be the independent, and the servant the dependent, self-consciousness. But in fact their relation is just the reverse. For it is the master, and not the servant, who remains dependent on the labor of another for his satisfaction in the world of things, and dependent on the servant self-consciousness itself for the one-sided recognition that is to confirm his self-certainty; and it is the servant, not the master, who manages to live without being dependent on the recognition of another, and who performs the acts of negation that render the objective world dependent on a self-consciousness, thus vindicating its self-certainty. Thus, as Hegel puts it, the truth of the master's self-certainty is found in the servant (PhG, ¶193). It requires only that the servant should become aware of this to obtain a new and higher kind of self-certainty, which is now confirmed as truth: in place of a one-sided "independence," its self-certainty takes the form of true "freedom" or "being with itself" (PhG, ¶197).

9.8 Freedom and Mutuality

In the *Phenomenology*, the master-servant dialectic results only in the free self-consciousness of the stoic sage—who, Hegel points out, is equally free whether, like Marcus Aurelius, he sits on the imperial throne or, like Epictetus, finds himself in the chains of a slave (PhG, ¶ 199). He does not here explicitly draw from the dialectic the Fichtean moral that recognition must be mutual, that it can be truly obtained only by those who give it in return, thus constituting a community of equals. But he does draw this conclusion in the *Encyclopedia's* account of this dialectic (EG, §§ 435–6). In both texts, the outcome of the dialectic is a "universal self-consciousness" that is truly free, because it stands in no need of the subordination of another.

Or, as Fichte puts it: "The only person who is truly free is the person who wishes to liberate everyone around him and who—by means of a certain influence [that of rational education and reciprocal communication]—really does so" (SW 6:309). The outcome of Hegel's indirect argument, therefore, is once again the same as that of Fichte's direct argument. Both yield the conclusion that the essential practical end of self-consciousness can be achieved only by standing in a

relation of mutuality with other self-consciousnesses, and that the practical task of self-consciousness is to act according to universal principles shared between it and these others. Hegel's transcendental method thus once again confirms in an original way the conclusions of Fichte's, and the two approaches are more continuous with each other than they are competing.

The common conclusion of both arguments is that free self-conscious beings, as a matter of transcendental necessity, require other free self-conscious beings in order to be, and that such beings must, on pain of inconsistency or incoherence, address each other fundamentally as equals. This is an important conclusion, which has profound and critical implications for all existing social orders. But it requires clarification, or even qualification in some important respects.

First, the equality between self-consciousnesses (or persons) holds only at a very abstract level: it does not require that people should, on pain of incoherence, receive equal amounts of anything. The conclusion of the Fichtean–Hegelian theory of recognition is consistent, up to a point, with the rejection of egalitarianism we will encounter in Marx in Chapter 11 (though not with Marx's wholesale repudiation of the standpoint of right, which we will also encounter there).

Second, it also does not require complete equality of recognition in regard to every way in which people regard each other, or value and esteem one another. People do require mutuality of a kind, even for forms of personal valuation and esteem where inequalities are unproblematic. Rousseau, as Frederick Neuhouser has read him, regards it as an essential, and also a morally permissible, human aspiration to seek the justified esteem of others for qualities that make us stand out as superior to others.

For example, if I wish to be esteemed by you as an expert on German philosophy in the eighteenth and nineteenth centuries, and to draw confirmation of my self-worth from your recognition of my merits, I do need to esteem you in return as a judge of these matters. I do not need to think you are my equal or my superior in this expertise (though I might rightly value your opinion all the more if I thought you were my superior). But it would be foolish, or even sycophantic, of me to base my self-valuation as a scholar on your esteem for my work if I thought you utterly ignorant of the history of modern philosophy and totally incompetent as a judge of philosophical arguments and theories. Conversely, you might draw self-esteem from the fact that I admire your performance as a baseball player; to do this, you do not have to consider me your equal in athletic skills, though you ought to esteem me as a judge of performance on the diamond. There is a kind of mutuality in relations of recognition pertaining to esteem, but these relations do not demand anything like equality.

Social relations involving only mutual respect for equal status would not, in Rousseau's view, be adequate to sustain the social relationships required for a

healthy kind of human self-worth. On this ground, Neuhouser takes issue with theories of recognition requiring absolute equality, including Hegel's theory (Neuhouser, 2008, Chapter 3, especially pp. 98–9). It is significant, however, that Fichte and Hegel are not concerned with equality as a condition for mutual recognition in general, but more specifically with the conditions for the recognition of persons as subjects of rights—external freedom, possessors of a rightful external sphere of free activity independent of domination by the will of another. Here the condition of equality does seem justified. To recognize another as a person I must regard the other as a free subject, entitled to a sphere of external freedom, personal independence, the conditions of a free life.

It is contrary to the conditions of recognition to hold, as some appear to do in the context of present-day discussions of social justice, that I count as a free person if I have the *opportunity* to earn the conditions of a free life, but am not *entitled* to those conditions. I have “earned” them, on this view, if I have enjoyed “success” in the “free market.” Otherwise, I may be in a position of economic dependency to those who have “earned” the wealth and power that enables them to dominate me. This, however, merely establishes on a “free market” basis the relation of lordship and bondage; it treats the conditions of free personhood as if they were something like the conditions of justified self-esteem—which do not require equality of rights or of mutual valuation. In this way, the extreme economic inequalities of capitalist society might be permitted to masquerade as conditions of universal freedom, when in fact (as is happening right now on an increasing scale) they merely reproduce the conditions of master and servant, condemning the majority increasingly to a condition of bondage.

Finally, as I hope is obvious, Fichtean and Hegelian arguments for equality of recognition do not preclude the possibility of there being outlooks, philosophies, or even entire social orders based on what Nietzsche liked to call a “rank ordering” of human beings: the premise that some are superior and others are inferior—the very principle that Hegel expresses in an abstract form as the “master–servant” relation. Indeed, this kind of inequality in regard to rights (in regard to the basic conditions of personal freedom) has in fact held in all civilized society (the fact that motivated Rousseau to write his *Discourse on the Origin of Inequality among Men*, and also Fichte to develop the foundations of his theory of natural right on the equal recognition involved in the relation of right). It is, as I have just suggested, part of the dominant ideology of our society today.

The arguments of Fichte and Hegel show only that all such outlooks or social orders are grounded at a very deep level on an inconsistency, a shared self-deception on the part of all who accept their premises. We know that many falsehoods, incoherences, and self-deceptions are powerful forces in human psychology and

human social life: powerful world religions and entire social structures are typically built upon them. The argument therefore gives us no strong grounds for hope that familiar forms of human inequality or oppression will soon disappear. At most, they show that inequalities of status between human beings are fundamentally at odds with human reason, and can be sustained only through the power of formidable, and all too familiar, irrationalities.

Hegel on Responsibility for Actions and Consequences

10.1 Hegel on Morality

When it comes to Hegel's philosophy, many serious errors and misleading half-truths have wide currency. Regarding moral and political philosophy, the most prevalent of these is probably still the false image of Hegel as "conservative, reactionary, quietist"; not far behind, however—and this is an error even more common among those who actually know something about Hegel's philosophy—is the idea that Hegel was an enemy of what he called "morality," and a proponent of "ethical life," which he saw as opposed to the "moral" standpoint. A related error is that "morality" for Hegel is only a nickname for Kantian ethics, so that Hegel's discussion of morality consists only in his critique of Kant. No doubt Hegel is critical of individualistic moral and social philosophies, such as Kant's, which he regards as proceeding in abstraction from the value of individuality in its social embodiment in modern ethical life, and "the moral standpoint" is a term that may refer to this philosophical one-sidedness (PR, §§ 33R, 135R). But in the *Philosophy of Right*, "Morality" is primarily a name for that distinctively modern way of thinking about the free will in which subjective freedom is its chief characteristic. "Morality" contains a theory of individual agency that differs significantly from Kant's. It is an alternative theory of moral action, which Hegel endorses, and it plays a significant and positive role in the practical philosophy he articulates in the *Philosophy of Right*.

In Hegel's ethics, the concrete human individual belongs to the ethical life (*Sittlichkeit*) of a rational community. But ethical life is a historical phenomenon, and a given form of ethical life is a practical expression of the understanding that spirit (or humanity) has of itself in a given age. In the modern age (Hegel thinks: the most rational society people have achieved, and the most rational they are as yet capable of comprehending), ethical life contains within it two important human self-conceptions, which are abstractions from ethical life, but necessary to its modern form. These are: (1) *abstract right*, in which human individuals are considered (in abstraction from their particularity) as *persons*, possessed of rights

and as owners of property (PR §§ 34–104); and (2) *morality*, in which individuals are considered as *subjects*, capable of governing their own lives and answerable to the demands of moral reason. Abstract right was a viewpoint that first developed in ancient Rome, corresponding to the conception of free status discussed in the previous chapter (PhG §§ 477–83; cf. §§ 197–9, with § 479, and EG §436). Morality, however, is a distinctively modern form of ethical self-consciousness (PhG §§ 596–8) corresponding to the distinctively modern institution of civil society (PR §§ 182, 206–7).

Subjectivity involves a conception of the self in which a universal will is distinguished from a particular will, and made normative for it. The task of the moral subject is to cultivate its particular will in such a way as to conform to the norms of the universal will. There is an obvious connection here to Kantian moral philosophy, but even more to the version of the same view put forward by Fichte in the *System of Ethics* (1798). These influences are the element of truth in the oversimplified and misleading claim that “morality” is Hegel’s nickname for Kantian ethics. But Hegel’s conception of moral subjectivity is a conception distinct from anything found in Kant or Fichte, and it is a self-conception that Hegel fully endorses. It is not merely a whipping boy for Hegelian criticisms, nor is it a conception to be rejected in favor of ethical life. Indeed, it is an essential moment of modern ethical life, and indispensable (in Hegel’s view) to the rationality of the modern world.

Within subjectivity, the distinctively moral conception of human beings, Hegel distinguishes three moments or aspects (PR §§ 109–13)

- (a) External objectivity (PR §§ 109–10);
- (b) Requirement: the relation of the universal to the individual (PR § 111);
- (c) Bearing on the will of others: the way (b) is manifest in (a) (PR § 112).

Moral action consists in the relation of these three moments (PR § 113). In other words, a moral subject is an individual who stands under rational requirements—moral duties (PR § 133). The subject manifests its subjective volition through external actions in the objective world that have a bearing on others—on their rights as persons, and on their welfare or happiness as subjects sharing the moral world (PR §§ 125–35).

We may therefore understand the unifying theme of Hegel’s presentation of the moral will as the ways in which the subjective will bears *responsibility* (in several different but related senses) for states of affairs that occur in the external world. Hegel takes it for granted throughout the *Philosophy of Right* that the human will is free. He agrees with Fichte that freedom is an essential determination of the will, and “unfree will” would be a contradiction in terms (PR § 4; cf. SL 4:159). But as in Fichte, the concept of free will itself plays only a marginal role in Hegel’s theory of

moral responsibility, because freedom of the will is seen as a condition of agency in general, not a special issue regarding moral responsibility. Therefore, in the remainder of this chapter we will have little or no opportunity to refer to the issue of free will. All that matters for moral responsibility is that the concept of moral subjectivity (just sketched above) should be applicable to human beings. Hegel does suppose that the concept of moral subjectivity applies to normal, healthy, human adults. Where it does not, or where it applies only imperfectly—as in the case of “children, imbeciles, lunatics, etc.” (PR, § 120R)—Hegel thinks that there is absent or diminished responsibility. Hegel takes it for granted that the application of the concept of moral subjectivity to people presupposes freedom of the will, but if someone were to think that this concept may be applied to people without this assumption, that person might employ Hegel’s theory of moral responsibility without ever needing to invoke the philosophical issue of free will. In Hegel as in Fichte, the issue of free will is only tangentially related to the topic of moral responsibility, and both topics should be regarded as mishandled if they are too closely linked.

10.2 Imputability in Kant and in Hegel

Hegel’s exposition of the theme of moral responsibility begins with his treatment of the traditional topic of *imputability* (*Zumutung*). This concerns the way in which actions and their consequences may be imputed (*zugemutet*) to moral subjects—in other words, the ways moral subjects may be held responsible for them: credited with them or blamed on account of them. A good point of departure for our understanding of Hegel’s treatment of moral subjectivity generally, and his account of moral action more specifically, is therefore his account of the way individuals may be held responsible for their deeds and for occurrences in the external world that are regarded as the results of these deeds. And a good initial point of reference for understanding this account is Kant’s treatment of the traditional topic of imputability in the *Metaphysics of Morals*. This will be useful because Kant presents a thoughtful and interesting treatment of imputability, but also because it will give us an opportunity to see right away that Hegel’s conception of morality differs in significant ways from Kant’s, and thus makes clear that “morality” in Hegel cannot possibly be merely a nickname for Kantian moral philosophy.

Kant follows tradition in calling an *action* (*Handlung*) a *deed* (*Tat, factum*) insofar as it is possibly imputed to an author (*Urheber, auctor*) and made the object of a judgment under laws by a suitable authority (*iudex sive forum*) and a ground for punishment or reward (*poena, praemium*) (MS 6:227). More specifically, what is imputed to an agent are the results or consequences of a deed. Kant proposes some very clear rules for it, based on whether the action is owed (required by a strict or

perfect duty), or wrongful (contrary to duty), or meritorious (pursuant to a wide or imperfect duty). The consequences of an action that is owed—both good and bad consequences—are not to be imputed to the agent. In the case of a wrongful action, all bad consequences are imputable, but no good consequences are imputable; and in the case of a meritorious action, all good consequences are imputable, but no bad ones are imputable (MS 6:228).

These principles play a role in Kant's well-known late essay about the right to lie, and its famous (or infamous) handling of the example of the murderer at the door. Assuming you are required to tell the truth to the murderer about his intended victim's whereabouts, and you comply with this duty, the death of the victim cannot be imputed to you. On the other hand, if you lie to the murderer (thereby doing wrong), and the victim (unbeknownst to you), has slipped away to a place where the murderer (believing your lie) encounters him and murders him, then the death can be imputed to you—as one of the bad consequences of your wrongful action (VRL 8:427; for Kant's views on lying, see above, Chapter 4, § 9, and also Wood, 2008, Chapter 14).

We may have a hard time swallowing the assumption that you are required by strict right to tell the truth to the murderer, but granted that assumption, Kant's conclusions about imputation seem reasonable ones. In general, Kant's principles of imputation give us quite reasonable results when applied to imputation under the law. Kant's principles do a very good job, for instance, of handling cases of liability to damages resulting from the deeds of agents that might affect the rights and claims of others. If my action is one that was owed (strictly required) under the law, and its omission not justified or excused by any special circumstances, then I should not be held liable for its consequences, however bad—whether these consequences were foreseeable or unforeseeable. It would be a sorry system of justice indeed that second-guessed those who have done what they were required to do, making them answerable for the bad results of actions the law required of them. If I take it upon myself to perform a wrongful action (and, once again, there is no justification or excuse for it, based on special circumstances), then it seems reasonable that I should bear the burden of all the bad consequences (whether foreseen or unforeseen), since those consequences would not have ensued if I had done what was required of me. Others had a right to rely on my doing what I should have: by doing wrong, I am the one who opened the flood-gates to whatever bad consequences resulted, and those who suffered damages should be able to recover from me the costs to them. If it is a question of reward, then none seems due to someone who did only what was required, and still less is reward due to someone who acted wrongly. Reward is due only to someone who acted meritoriously.

Kant's principles are clearly intended, however, to apply not only to issues of right, but also to the appraisal of deeds (*imputatio diiudicatoria*) from a merely ethical or moral point of view (MS 6:227). And here their results seem far more questionable in many cases. They fail to countenance judgments of imputation (both positive and negative) of which we feel fairly confident, and they are also insufficiently appreciative of situations of moral ambiguity. Here I am not thinking mainly of wrong actions that may have good results, for which the agent might hope to take credit—for it seems quite defensible (and Hegel will be seen wholly to agree) that people who act immorally should not be able to take credit for the contingent good consequences of their bad behavior. More serious problems arise concerning owed or meritorious actions whose consequences are mixed or ambivalent. If I perform an action that was strictly owed, but whose consequences turn out to be very beneficial, I may choose to deflect praise by saying that "I was only doing my duty;" but we may feel this would be false modesty. On the other hand, if I perform an action that is owed, or even meritorious, perhaps because of its consequences for one party, but which also (without losing its status as required or meritorious on that account) does some harm to another party (perhaps even a harm I foresee with regret), I don't think we are comfortable saying that this harm was not imputable to me at all, simply on the ground that the action was owed or meritorious. We tend to think that the harm is still in some way imputable to me—perhaps that I even owe those who suffer this harm some recompense for it, or at least an apology acknowledging my responsibility for it (even if I can also rightly say that, since the action was owed by a strict duty, I "had no choice" but to do it). The moral ambiguity present in such cases is not properly acknowledged in Kant's account. We will find that Hegel's theory of imputation possesses a degree of subtlety in this respect that Kant's does not, and that its accommodation of moral ambiguity is one of its virtues.

There is an even more basic difficulty with Kant's views here, however, that will exhibit an even sharper contrast with Hegel's theory of imputation. This is that apart from specific legal provisions, dealing with liability to damages for the consequences of actions, the whole topic of imputation of consequences might seem to look like an insignificant afterthought in the context of Kantian ethics. This is because for Kant the basic moral judgment of an action is apparently independent of its consequences. Kant famously holds that the unlimited goodness of the good will is entirely independent of its good (or bad) results—so that the good will "shines like a jewel for itself, as having its full worth in itself. Utility or fruitlessness can neither add to nor diminish its worth" (G 4:394). Moral judgment is about the volition (good or bad) that is exhibited in the action, which is to be gauged by the agent's maxim in performing it, and not at all by its results. An action may be

judged right or wrong on account of the agent's *end* in performing it (since this is often part of the maxim, and the agent's end goes to the goodness or badness of the agent's will). If the action is successful, the accomplishment of its end will be among its results. But for Kant those results matter only because they belong to the agent's maxim, or goodness of will, and not because they occur in the world as something whose imputation to the agent might make a difference independently of our judgments about the agent's will.

Further, for Kant, whether an action, is owed, or wrongful, or meritorious, is to be judged according to its maxim, and the imputation of its consequences is decisively shaped by this. It is difficult to see why, outside legal contexts (where civil damages or the degree of gravity of a crime might be at stake), a Kantian agent should be interested at all in the imputability of consequences. All that really seems to matter is the goodness or badness of the volition contained in the action (in its maxim). If I have a good will, then of course *I care* about the good or bad consequences (the utility or fruitlessness) of my good actions, because this caring is part of what it is to set the good ends that go with having a good will. But why should I, or anyone, be concerned about whether the fortunate or unfortunate consequences of my actions are *imputable* to me? My purely moral evaluation of myself, before the inner forum of conscience, seems to be exhausted by the question whether my will is good. If in fact we do consider the consequences of our actions relevant to their moral evaluation, over and above the evaluation of the good or bad volition that led to them (and this alone is what a theory of imputation of consequences would be for), Kant has a hard time saying why we should.

Hegel, by contrast, regards the moral will as primarily a relation to external objectivity, and its value as a moral will depends on whether and how what happens in the world can be imputed to it. We can see this difference clearly in the differing roles played by the concepts of "action" and "deed" in the action theory of the two philosophers. For Kant, an *action* (*Handlung*) is an expression of volition (of a maxim); a *deed* (*Tat*) is an action insofar as its consequences fall under principles of imputability to the agent (MS 6:227). For Hegel, by contrast, a *deed* "posits an alteration to the given existence (*Dasein*)" (PR § 115); and a deed is considered an "action" insofar as it is imputable to the will of the agent (PR § 117). In other words, for Kant, what is primary is volition, which is taken as having a reality prior to and independently of the changes it may bring about in the world; this is the basis of actions, while deeds are actions considered in a certain way—in terms of the imputability of their external consequences. For Hegel, however, there is no morally significant volition apart from deeds; *moral* volition or action is merely the way we consider deeds in relation to the agent, as a volitional moral subject. For Hegel, in fact, the moral will itself is constituted by the way deeds and consequences may be imputed to the subject.

10.3 Responsibility, Purpose and Intention

The starting point of moral action for Hegel is the *deed*—an alteration in the objective world that is brought about in some way by the moral subject. The most abstract and general relation of the subject to its deed is that of *being responsible for it* (*daran Schuld sein*). Despite the possible moral connotations of the word *Schuld* (which in German can mean either “debt” or “guilt”), what Hegel means by “responsibility” in this sense is a merely causal relation, entailing that the action may have the abstract predicate “mine” applied to it, but not implying any moral imputability whatever (PR § 115). According to Hegel, the causes of a historical event, which may be manifold, most of them having no moral agency at all, are “responsible for it” in this sense. Hegel thinks of “responsibility” as well suited to the abstract understanding, which can view any event in a number of different ways and treat any of a number of factors as “its cause”:

Every individual moment which is shown to have been a condition, ground or cause of some such circumstance and has thereby contributed its share to it may be regarded as being wholly or at least partly responsible for it. In the case of a complex event (such as the French Revolution), the formal understanding can therefore choose which of a countless number of circumstances it wishes to make responsible for the event (PR § 115R).

Moral imputability for Hegel turns not on “responsibility” (*Schuld sein daran*), but rather on the application of two other concepts, which Hegel designates by the terms “purpose” (*Vorsatz*) and “intention” (*Absicht*).

Responsibility for an action belongs to the subjective *will* only for “those aspects of its *deed* which it knew to be presupposed within its end, and which were present in its *purpose*.” This Hegel calls the “right of knowledge” (PR § 117). The “purpose” apparently encompasses everything the agent was aware would happen, and not only those aspects the agent specifically desired to bring about. But it excludes aspects the agent had no way of knowing about, and also remote consequences of the action, which might be brought about by “external forces which attach it to things quite different from what it is for itself, and impel it to remote and alien consequences. The will thus has the right to accept responsibility only for the first set of consequences, since they alone were part of its *purpose*” (PR § 118). As an example, Hegel cites Oedipus’s parricide as something not contained in his purpose, when he fought with and killed the old man he met at the crossroads. Hegel thinks it was part of the “solidity” or “noble simplicity” (*Gediegenheit*) of the ancient conception of action that it did not recognize this subjective “right of knowledge” as we do in the modern world (PR § 117A). This is an important aspect of the way in which “morality” is a distinctive characteristic of modern ethical life that was lacking in ancient ethical life.

The “purpose” of a deed, however, includes not only what the agent specifically intended or took as an end, but also what the agent knew would occur, even if it was not desired. If a pilot bombing an enemy military installation knows that the school next door is also going to be destroyed (killing many innocent children), the death of the children, and not only the destruction of the enemy installation, is part of his purpose. The purpose of my deed may include even consequences I did not know about, if they belong to “the nature of the action itself” (PR § 118R). In general, for Hegel, the “nature” of anything includes what we would grasp about it from rational reflection on it and its connection with other things (EL § 23). In the case of a deed, these include connections with its consequences (PR § 118; VPR19: 94). Consequently, the nature of an action includes all the consequences that would be known by rational reflection: “In general it is important to think about the consequences of an action because in this way one does not stop with the immediate standpoint but goes beyond it. Through a many-sided consideration of the action, one will be led to the nature of the action” (NP 230).

Hegel’s talk about the nature of an action may be recognizable as more familiar to us if we see how it relates to our judgments about people when we say that they did (or did not) *know what they were doing* (in the sense of fully *realizing* or *appreciating* what they were doing). A mature adult is responsible for knowing what will, or may, result from her actions (for instance, from starting a fire; or for a president: starting a war, authorizing torture, deregulating the banking system). Not to know this is perhaps a cognitive or intellectual defect, but it is a ground for moral criticism and a basis for holding someone responsible for the results of what they do, even when it differs from what they hoped or expected.

This is how Hegel intends to deal with the example of an arsonist who sets fire only to one house (or one stick of furniture), but ends up destroying a whole neighborhood (PR § 119). Hegel holds that the fully developed consequences of the action belong to the arsonist’s purpose, because it belongs to the very nature of the act of setting a fire that it may spread out of control: “The *dolus directus* or direct purpose is, for example, setting fire to the first piece of wood, and the *dolus indirectus* contains all the further consequences. These belong to the nature of the action itself, which posits their possibility along with it. The man must know this” (VPR 4:326; cf. EG § 505). Some views might hold that the agent is to blame in this case not for the spread of the fire beyond what he directly intended, but only for being insufficiently attentive in thinking about how far the fire might spread. Those views might fault him for not knowing what he should have known, but they cannot impute to him the destruction of the entire neighborhood. Hegel, however, by including the entire nature of the action within the agent’s purpose, is making the arsonist responsible directly and originally for all the possible

consequences that he should have reckoned with, because these belong to the nature of his action.

The “purpose” of an action for Hegel marks out *what* the subjective will is responsible for, or *what* objective occurrences may be imputed to it. But it does not tell us everything we need to know about the action in order to judge the agent’s subjectivity from the standpoint of morality. In order to make moral judgments, we must also take into account how the agent’s *subjectivity* was related to these occurrences. This is comprised within what Hegel calls the agent’s “intention” (*Absicht*). The *purpose* of the bomber pilot, in the above example, includes both the destruction of the enemy installation and the death of the children in the school. But as a moral subject he is related to these two results in very different ways. The first was his direct aim, the second only a regrettable consequence of the way he had to carry it out under the circumstances.

As a subject, says Hegel, I am a thinker, and bring my actions and their consequences under a “universal.” This is what Hegel calls the action’s “intention.” But by a “universal” here Hegel does not mean merely some general description under which the action might fall (such as “burning” or “killing”), but the consequences organized into a complex by the agent’s thought: this is a general point for Hegel about the philosophical meaning of words like “universal” and “concept,” and the “concreteness” they involve (see EL §§ 9, 160, 164, R, 176–7, R). In this case, the “universal” under which the agent brings the action and its consequences implies a determinate act of abstraction, which Hegel connects to the etymology of *Absicht*—“looking away.” It refers to the determinate thought the agent has about a particular deed and the concrete complex of actions and consequences it involves—especially that abstract aspect of this complex that constitutes for the agent its “subjective essence” (PR § 121A).

10.4 Intention and Motive

In a non-Hegelian philosophical jargon, we might call the “intention” of an action the “desirability characterization” it would have for the agent, or the “description under which” the agent intended to perform it. (For an extensive and perceptive attempt to relate Hegel’s theory of action in this section to more recent “analytical” approaches to action theory and philosophy of mind, see Quante, 2004. The present discussion is not the place, however, for an evaluation of Quante’s project or a general assessment of the appropriateness and limits of interpreting Hegel through the lens of analytical action theory.) “The *right of intention* is that the *universal* quality of the action shall have being not only in itself, but shall be *known* by the agent and thus have been present all along in his subjective will” (PR § 120). My action should

be judged according to the universals under which I know and will it, and this judgment should take account of the descriptions under which the action and its consequences were desired by me in performing it. In morally assessing my action, different aspects of the action, belonging to its purpose, should be distinguished from one another depending on whether they were willed by me as an end, or as a means, or as a regrettable side-effect. The pilot “intends,” in this narrow sense, to destroy the enemy installation, but not to kill the children. The arsonist “intends” to avenge himself by setting fire to his neighbor’s antique chair, but not to destroy the whole neighborhood. An agent, as a moral subject, is to be held responsible for his deed only by taking into account the way he thought about it.

Notice that an “intention” for Hegel is always *the intention of a deed*—the way the agent has thought about some actual alteration in the objective world. There are no morally significant intentions independently of, or in abstraction from, actual deeds in the world. This is one of the several meanings of Hegel’s slogan-like pronouncement: “The truth of the intention is only the deed itself” (PhG ¶ 159). Hegelian ethics, unlike Kantian ethics, does not judge “intentions” (or “maxims”) apart from the deeds that embody them. The moral subject is always an agent in the objective world, not a subject of mere “volitions” that are good or evil (“having their full value in themselves”), irrespective of whether they may or may not have any results. This is related to the separate point that Hegelian intentions are also not subjective or mental *causes* of deeds or occurrences. Instead, an intention is the way I *think about* what I do in doing it, and morality should treat the thought and doing as mutually necessary to each other.

There is no word in German that answers precisely to the meaning of the English word “intention.” But Hegel’s terms *Vorsatz* and *Absicht* might both be loosely translated with that English word, and I think Hegel’s distinction between “purpose” and “intention” is quite close to a distinction that T. M. Scanlon has recently drawn between two ways in which we speak about the intention of an action.

“Intention” is commonly used in a wider and a narrower sense. When we say that a person did something intentionally, one thing we may mean is simply that it was something that he or she was aware of doing or realized would be a consequence of his or her action. This is the sense of “intentionally” that is opposed to “unintentionally”: to say that you did something unintentionally is to claim that it was something you did not realize you were doing. But we also use “intention” in a narrower sense. To ask a person what her intention was in doing a certain thing is to ask her what her aim was in doing it, and what plan guided her action—how she saw the action as promoting her objective. To ask this is in part to ask what her reasons were for acting in such a way—which of the various features of what she realized she was doing were features she took to count in favor of acting in this way. (Scanlon, 2008, p. 10).

Scanlon’s wider sense of “intention” corresponds to what Hegel means by *Vorsatz*; his narrower sense, to what Hegel means by *Absicht*. To say that parricide was not

included in Oedipus's *Vorsatz* (in killing the querulous old man at the crossroads) is to say that he did not realize (and could not have known) that he was killing his father, so that his act under the description "parricide" was unintentional. To identify the *Absicht* of an action, the universal under which the agent brings it in acting, is to identify the agent's reasons for doing it, or the features of the action (or its results) that counted in favor of acting that way. If there is a difference between Hegel's distinction and Scanlon's, I think it has to do with the way Hegel means us to consider (for the purpose of assessing an agent's responsibility for deeds and their consequences) some of the ways in which moral subjects are accountable for things they should have known, and should have thought of, even though they did not. Hegel's theory of morality in fact makes quite strong demands on agents in this respect.

The attempt is often made to understand both intentions and motives as kinds of mental *causes* of actions. Hegel rejects this. It seems to belong to what I have already (in Chapter 8) characterized as the Cartesian conception of mind, which seeks to understand mental states, including actions, as "inner"—whether the metaphorical inwardness of the mental *poêle* in which Descartes meditates or the literal inwardness of brain processes. Hegel understands both intentions and motives as ways the agent thinks about the action. A motive (*Beweggrund*), for Hegel, is only "the particular side of the intention" (PR § 121A); that is, the aspect of the intention that represents the particular agent's subjective satisfaction (PR § 121).

Hegel's account here seems to me rather close to that of Elizabeth Anscombe, when she writes: "To explain one's own actions by an account indicating a motive is to put them in a certain light." (Anscombe, 1963, p. 21). Neither motives nor intentions should be regarded as *causes* of actions. (Actions are grounded on reasons; they are not correctly thought of as the effects of causes; here is a point at which Hegel's conception of moral subjectivity might indeed touch on some of the issues surrounding the problem of free will.) Intentions and motives are instead ways that embodied agents in a social world conceive of those events to which they apply the concept of moral subjectivity. Those natural sciences that treat matters causally may teach us a lot about the physiology of human beings, but human *actions* will forever elude their grasp (and therefore actions, intentions and motives will be either misconceived or else dismissed as illusory by the scientific philosophers who see things only in these causal terms).

10.5 The Right of Objectivity, and Negligence

"To attempt to justify something in terms of its intention is to isolate an individual aspect completely and to maintain that it is the subjective essence of the action" (PR § 119R). The attempt, however, in Hegel's view (as this quotation perhaps already implies), is not always successful, and may often involve deception

(or self-deception). This point is due to what Hegel considers the essential complement to the “right of intention,” namely “to what we may call the *right of objectivity* of the action”—“the right of the action to assert itself as known and willed by the subject as a *thinking agent*” (PR § 120). The right of intention, in other words, must not be seen as a right on the part of the agent to have an action judged solely on the basis of the agent’s own (perhaps one-sided and self-serving) representation of it. The bomber pilot, in our earlier example, has the right to have his action judged by its aim of destroying the enemy installation, with the recognition that killing the children was something he did only with regret. But he also cannot treat the death of the children as something for which he bears no moral responsibility at all (as Oedipus, in Hegel’s view, bears none for killing his father or conceiving children by his mother, which belonged neither to his intention nor his purpose). The death of the children belongs to the bomber pilot’s “purpose,” even if not to his “intention” (though we will see presently that it may belong to this as well).

The moral agent is a thinking being, whose intentions are to be judged by objective standards of thought. This is related to the point that intention, as well as purpose, takes account of the “nature” of the action—the “external connections inherent in its nature . . . Hence in murder, it is not a piece of flesh which is injured, but the life itself within it” (PR § 119R). If I injure a person’s body in a way that rational reflection would recognize as endangering their life, I may not disclaim responsibility for their death by saying that I intended only to injure their flesh and not to take their life. Hegel realizes, of course, that there might be difficult questions here. It is often possible in fact to intend only to punch someone in the jaw, and not realize that you might actually be risking killing them: “It is the nature of the finite deed itself to contain such separable contingencies.” Where Hegel sees such ambiguities as coming to an end is with the agent’s awareness that an action is wrong according to objective standards. If an action that contributes to the agent’s happiness happens also to be wrong, a thinking agent is accountable for knowing this, and his intention must be thought of as including not only “promoting my happiness,” but also “doing something wrong.” Accordingly, the right of intention is seen by Hegel as precluding or diminishing responsibility for wrongdoing only in the case of “children, imbeciles, lunatics, etc.” (PR § 120R). The thinking agent is responsible not only for what she actually thinks is right and wrong, good and bad, but also what she *has reason to think*:

The right of the subjective will is that whatever it is to recognize as valid should be perceived by it as good, and that it should be held responsible for an action—as its aim translated into external objectivity—as right or wrong, good or evil, legal or illegal, according to its cognizance [*Kenntnis*] of the value which that action has in this objectivity (PR § 132).

One can be *cognizant* of many things that one does not actually know or realize. If I have every reason to know that what I am doing is wrong, evil, or illegal, then I can be judged *cognizant* of that. I cannot disclaim responsibility for my wrong act on the ground that “I did not know it was wrong.” I cannot demand that my act be imputed to me *only* under the intention “that it was something *good*” (e.g. good for me, as satisfying my momentary passion for revenge, or good in whatever way it now pleases me to rationalize it). “Just as what the arsonist sets fire to is not the isolated area of wood an inch wide which to which he applies the flame, but the universal within it—i.e. the entire house—so too is the arsonist himself, as a subject, not just the individual aspect of *this* moment or this isolated passion for revenge” (PR § 132R). On similar grounds, I think we must say of the bomber pilot that the death of the children belongs not only to his *purpose*, but also in a way to his *intention*, even if it does not belong to his *aims*, narrowly considered. It certainly belongs to his intention if we assume that it is *wrong* of him (e.g. contrary to the right of war) to kill innocent civilians, and that he is cognizant both of the fact that he is killing innocent civilians and that this is contrary to the right of war. The intention in light of which an act is morally judged, in other words, must take account of the agent’s own thoughts about the action, but it will not be limited to them if they do not include everything relevant to what a thinking agent would know about the nature of the action or its value according to the objective standards of right, morality, or ethics.

Because Hegel’s account focuses attention chiefly on what the agent knew about the action and its consequences, and on the agent’s intention in performing the action, it might be doubted that he can give an adequate account of our responsibility for consequences that result from carelessness, recklessness, or negligence. This is a charge made by Karl-Heinz Ilting, who thinks Hegel’s account cannot deal with negligence, or the distinction between foreseen and unforeseen consequences of actions (Ilting, VPR17, n 125, pp. 303–4). But the considerations we have been examining provide us with a compelling and cogent answer to the charge. Hegel considers both the purpose and the intention of an action to include its *nature*—the complex of its consequences that would have been brought to light by rational reflection—and also the *right of objectivity*—the action’s intention must include its relation to laws and ethical duties. Therefore, these aspects of the action, and these consequences, are imputable even if the agent is (negligently) unaware of them.

Perhaps one might worry instead about whether Hegel’s theory can distinguish intentional wrong from negligence, since he imputes the nature and ethical objectivity of an action to the agent irrespective of whether the agent is actually aware

of them (as I did, in Wood, 1990, pp. 143–4). But there should be no problem here either, as long as we can take account, in each case, of *why* a given consequence or objective ethical determination is included in the action's purpose and intention—whether it is there because it was actually known and willed by the agent, or whether it is there because it belongs to the objectivity and nature of the action, although the agent was thoughtlessly or carelessly unaware of this.

Someone might still worry that Hegel's account—by making us responsible for negligent actions through inclusion of the relevant aspects of the action and its consequences in its purpose and intention—cannot explain the general moral fact that harm caused by negligent actions is considered less blameworthy than the same harm caused intentionally. For example, Hegel holds that someone firing a gun into the woods, intending to hit an animal but killing a human being, is guilty of manslaughter but not of murder (VPR 2:423, 3:358; VPR17:78). We might still wonder how Hegel can justify treating the negligent causing of a death as less blameworthy than the intentional causing of a death. But it seems to me that this too is a matter that he can say ought to be settled on a case-by-case basis. It is surely true in the above example that the careless hunter is less blameworthy than a deliberate murderer would be.

Perhaps what we want is a justification for some supposedly *general truth* that wrong actions done from negligence are always less culpable than wrong actions done from malice. But I question whether this is a truth at all. For Hegel, the moral subject is a *thinker*, and moral conduct is always to be measured by diligent adherence to the standards of rational thought, and never merely by some sentimentalist conception of “goodness of will” based on non-rational feelings. Nor are rational standards applicable only to the deliberate volitions of the subject, as a Kantian might suppose. Goodness or badness of will is one kind of defect in a moral subject, but so are carelessness, negligence, and irresponsibility. And which defect is the more blameworthy may depend on the details of the case. Sometimes we see government officials (or academic administrators) do harm to an individual or group, and it is unclear whether they do so out of hostility to those they harm or out of mere negligence in performance of their duties. No doubt such administrators act wrongly if they treat people maliciously, but there might sometimes be mitigating conditions making their malice understandable, whereas diligence and care in the exercise of power may be an even higher requirement on them, whose absence is quite inexcusable and an even more serious moral failure than the presence of ill will. I have seen some such cases in the academic world (and I would bet you have too) where the degree of carelessness and irresponsibility required to explain some official's abuse of power would have been so extreme and outrageous that a more charitable interpretation of their behavior is that they did the harm out of ill will.

10.6 “Double Effect”

Hegel’s theory of imputability might be usefully compared at this point with the so-called doctrine of “double effect” employed in Roman Catholic theology. (For a good critical discussion of the doctrine of double effect as it is usually now interpreted, see MacIntyre, 2011.) This doctrine seems to be based on Thomas Aquinas’s discussion of killing in self-defense:

Nothing hinders one act from having two effects, only one of which is intended, while the other is beside the intention.... Accordingly, the act of self-defense may have two effects: one, the saving of one’s life; the other, the slaying of the aggressor... Therefore, this act, since one’s intention is to save one’s own life, is not unlawful, seeing that it is natural to everything to keep itself in being as far as possible (Aquinas, 1988; “On Killing,” drawn from *Summa Theologica* II-II, Q. 64, art. 7).

According to its standard interpreters, the doctrine of double effect holds that an agent is released from responsibility for a “second effect” of an action under the following conditions:

- “1. The act itself must be morally good or at least indifferent.
2. The agent may not positively will the bad effect but may permit it. If he could attain the good effect without the bad effect he should do so. The bad effect is sometimes said to be indirectly voluntary.
3. The good effect must flow from the action at least as immediately (in the order of causality, though not necessarily in the order of time) as the bad effect. In other words the good effect must be produced directly by the action, not by the bad effect. Otherwise the agent would be using a bad means to a good end, which is never allowed.
4. The good effect must be sufficiently desirable to compensate for the allowing of the bad effect.” (Connell, 1967, p. 1021)

Hegel’s theory of imputability would take issue with this doctrine, I believe, on several points. Let us consider the Hegelian position on each of the above claims:

(1) For Hegel, whether a deed is morally imputable to an agent is not necessarily dependent on a judgment that the act is good or bad. The universals under which the intention an agent brings an action may include those of right, morality, or ethical duty, but they also may not. And the judgment we make about the moral or ethical meaning of the action is in any case subsequent to determining imputability, not prior to it, as it always is Kant’s theory of imputability, which always makes the imputability of good or bad consequences of an action dependent on whether an action conforms to a strict duty, violates a strict duty, or fulfills a meritorious duty. By contrast, Hegel holds that we must judge agents based solely on

their actions (PR § 124), and the morality of actions based at least in part on their imputable consequences (in other words, what belongs to the agent's purpose and intention). Some actions (for example, the violation of the rights of a person, or of a moral or an ethical duty) would be wrong no matter what the imputable consequences. But Hegel thinks that our judgment of an act's the goodness or badness itself should depend on the goodness or badness of its imputable consequences, so in Hegel's view, (1) would make into a condition of imputability a judgment that might only be made after issues of imputability have first been settled. Also, Hegel would reject (1) because he thinks that an action that is not wrong—even an action that fulfills a duty—might have bad consequences that are imputable to the agent, if they belong to the agent's purpose and intention.

(2) For Hegel, what the agent can be said to will, and how the agent can be said to will it, is precisely what judgments of imputability are about. We cannot determine what the agent wills independently of what is imputable to the agent, and make the imputability of a consequence depend on this, as (2) apparently proposes to do. For example, the pilot in the above example might intend the destruction of the enemy installation under the concept: "primary objective of my mission," but if the death of the children belongs to the nature of the action (if their death is among the results that rational reflection would have understood as connected to the action), then this too belongs in some way to the pilot's intention (perhaps as "collateral damage," or "regrettable side effect"). The arsonist intends to set alight a single piece of furniture, and this belongs to his intention as "primary aim." But if the possible destruction of the entire neighborhood belongs to the *nature* of the action, it too must be included under some description in the arsonist's intention.

(3) would appear to be motivated by the scriptural principle that we must not do evil that good may come (*Romans* 3:8). Hegel has no reason to take issue with this principle. He emphatically rejects as corrupt the maxim "the end justifies the means" (PR § 140, R, A), and he understands certain actions as by their nature wrong (for instance, violations of abstract right or valid ethical principles), and therefore incapable of being justified by the supposed good intentions behind them (PR § 126). So an action that is wrong may not be justified because it is a means to something good. Something bad may, however, be part of the means to something good. A bad consequence, if imputable to the agent's deed, need not be regarded by the agent as a *means to good* in order to be imputed to the agent, and the deed judged on the basis of it.

(4) It is over this claim, it seems to me, that there is the deepest disagreement between Hegel's theory of responsibility and the doctrine of double effect. The doctrine of double effect seems to treat a bad consequence, if it is the "secondary" consequence of an action, just as if this secondary consequence were not

imputable to the agent at all—as if it were no more imputable than an unfortunate result of an action that the agent could not have foreseen. Even in the case of killing in self-defense, however, this is plainly untrue. If I kill an assailant in self-defense, I may well do so judging that the only way to save my life is to end the assailant's life. In that case, my act of killing is clearly intentional.

On this point, the doctrine of double effect seems to me dubious Thomism. When Aquinas says that in the case of self-defense the killing of my assailant is “beside the intention,” it is not evident that we must interpret him as saying that it is *unintended*; we could just as well take him to mean only that it is not the *primary* intention, or is not intended *as an end or for its own sake*. (This, in fact, may be what Connell means in (2) when he says the bad effect may be “indirectly voluntary;” but this does not tell us whether the effect is *intended*.) Some interpreters of the doctrine of double effect do explicitly say that for the doctrine to apply, the bad effect must be *unintended* (Mangan, 1949, p. 43). However, if that were accepted, then it seems unlikely that the doctrine could be correctly applied even to the supposedly paradigm cases.

Hegel would hold that any effect of my action that falls within my cognizance or belongs to the nature of the action must also belong to my intention, under some concept or other. If it belongs to the nature of my act of self-defense that I take my assailant's life, then his death belongs to my intention as well as my purpose, and it is imputable to me, even if my action is considered justified and not wrong. The attacker's death may be intended only as a means to my rightful defense of my life, rather than as an end in its own right. But I cannot honestly claim that it is unintended.

Hegel also need not accept the idea that whether the action is wrong depends, as in (4), on whether the value of the primary intended effect outweighs the value of the secondary effect. I need not judge that the preservation of my life has greater value than the life of my assailant in order to be justified in defending myself. Whether it is justified depends, rather, on whether the act of self-defense falls within my right. And even a justified action may be the basis for imputing bad consequences to the agent. I may be justified in killing my assailant in self-defense, but I did nevertheless bring about a death, and this death is still imputable to me, as part of my purpose and my intention, even if I committed no crime, and incurred no blame.

It is sometimes claimed that the doctrine of double effect is heterogeneous in intent and motivation, or seeks to reconcile principles that are harder to unite than its defenders want to admit (see MacIntyre, 2011). For instance, the doctrine may be seeking to reconcile such actions as sacrificing one's own life or killing an enemy

in war with the principle that it is wrong to kill, by treating the death as something for which the agent is not responsible. But from a Hegelian standpoint, this seems dishonest. There is no question that a death is imputable to the agent in these cases, even if the action that brings it about is justified. It ought to offend our moral sensibilities if someone tries to pretend otherwise.

The doctrine of double effect seems to be aimed at eliminating the moral ambiguity involved in actions that we think are on the whole good, or at least not evil, but which foreseeably do bring about harm or bad consequences. The doctrine wants to treat such harm and bad consequences as if the agent were *in no way* responsible for them. But honest moral agents know that they bear a moral burden in such cases, and that a profound moral ambiguity often attaches even to actions that we think are on the whole good. If the doctrine of double effect is intended to deny this, then it is profoundly mistaken. Hegel's theory, by contrast, does not flee from moral ambiguity, but embraces, accepts, wrestles with, and tries to understand it.

10.7 "Moral Luck"

One set of issues about moral responsibility for consequences arises in connection with what some philosophers call "moral luck," and especially about what Thomas Nagel has called "resultant luck"; that is, the fact that two agents, performing identical actions, may apparently incur very different degrees of praise or blame, depending on differences in the consequences of their actions that do not seem to be up to them at all, but are due to good or bad fortune (see Nagel, 1993, p. 60). Hegel's views on moral responsibility have, I believe, some original insights to contribute here.

First, because Hegel's account begins not (like Kant's) with abstract, subjective "volitions" but with actual deeds, his view challenges the facile assumption that we should ever judge two agents to have performed actions that are truly "identical," even though they have very different consequences. This assumption, Hegel thinks, is based on the notion that one can isolate the subjectivity or "inner" aspect of an action distinct from its "outer" aspect, and equate two "subjectivities" in cases where their "outer" deeds are very different. Hegel's rejection of that notion is blunt and emphatic:

What the subject is, is the series of its actions. If these are a series of worthless productions, then the subjectivity of volition is likewise worthless; and conversely, if the series of an individual's deeds are of a substantial nature, then so is his inner will (PR § 124).

What a human being does should be considered not in its immediacy, but only by means of his inwardness and as a manifestation of that inwardness. But with that thought we must

not overlook the point that the essence and also the inward only prove themselves as such by stepping forth into appearance. On the other hand, the appeal that human beings make to inwardness as an essence distinct from the content of their deeds often has the intention of validating their mere subjectivity and in this way of escaping what is valid in and for itself (EL § 112A; cf. EL § 140).

There is no absolute point of comparison, then, between two people whose deeds have had very different consequences, that might enable us to say that they were “inwardly the same” (hence in some sense “morally equal”) and that the different consequences of what they did were due only to factors “external” to their moral subjectivity (and hence to “mere luck”). The doctrine presented in the above quotations might be taken as a second distinct meaning to Hegel’s slogan-like pronouncement: “The truth of the intention is only the deed itself.”

In the depiction of two agents whose “inner volition” is supposed to be the same, but whose deeds have very different consequences, there is a remnant of that Cartesian conception of mind—in this case, of action—which seeks to locate and isolate thoughts and even deeds somewhere “within” the whole human agent, regarded as an embodied and a social being (whether this “within” belongs to an immaterial substance or to a human brain within the skull). Fichte’s theory of mind, as we saw in Chapter 8, seeks to overturn this conception of mind, and especially of action. Fichte denies there are any purely immanent actions: every act of mine is an action of my body on the external world (SL 4:89–93). Hegel clearly follows Fichte in this respect, and even develops the anti-Cartesianism further. The moral subject is nothing but its actions, inner volition is to be judged only by its outer expression (PR § 124).

Hegel clearly does not hold, of course, that if an action is clearly wrong, violating morality or ethics, it can be justified merely because it happened to have good consequences. Rather, his thesis is that the inner and outer aspects of an action must be considered together and evaluated together, according to their systematic meaning. Suppose, for instance, that the social act of providing alms to the poor through private, voluntary (e.g. “faith-based”) charity has the systematic effect of keeping the poor wretched and oppressed. This systematic result should lead us to revise (downward) our estimate of the moral value of the inner subjective disposition to voluntary religious almsgiving. On the other hand, suppose Hegel is right that true beneficence, the beneficence that truly helps those in need, is found only in “the intelligent universal action of the state” (PhG ¶ 425). In that case, the political disposition to replace faith-based voluntary charitable giving with state-administered aid to the poor would be the inner disposition that is truly to be admired and esteemed from a moral standpoint. Such judgments, of course, have to be made from a systematic consideration of the social meaning of these inner

dispositions, not merely on the basis of the accidental consequences of their manifestation in some individual case.

Hegel does not deny, however, that it makes sense to say of what an agent did that its consequences might have been very different from what they were, and that these differences might have been due to contingencies, perhaps unforeseen and unforeseeable, that were beyond the agent's control. Here too he has reference to his example of the arsonist:

It is certainly the case that a greater or lesser number of circumstances may intervene in the course of an action. In a case of arson, for example, the fire may not take hold, or conversely, it may spread further than the culprit intended. Nevertheless, no distinction should be made here between good and ill fortune, for in their actions, human beings are necessarily involved in externality. An old proverb rightly says: "The stone belongs to the devil when it leaves the hand that threw it." By acting, I expose myself to misfortune, which accordingly has a right over me and is an existence of my own volition (PR § 119A).

When Hegel says here that "no distinction should be made here between good and ill fortune," he means that I cannot offer my ill fortune as any *excuse* for the harm I have caused (or for the failure of my fruitless attempts to do good), by comparing the actual consequences of what I did with some different and better outcome that I claim would have occurred if only I had been luckier. In saying that when I act, I expose myself to misfortune, and that it "has a right over me," Hegel is using the term "right" (as he often does) both in an everyday sense and also in his technical philosophical sense, in which "right" means "any existence in general which is the existence of the *free will*" (PR § 29). In other words, he is saying that in acting, I freely choose to expose myself to misfortune, and this is what gives it a right over me, so that I am responsible for its results.

An interesting—and I believe insightful—aspect of Hegel's theory of imputability here is that he treats the imputability of deeds and consequences (our "responsibility" for them in the broadest sense) as part of our conception of moral agency ("moral subjectivity") that is prior to our judgments about or attitudes towards what we have done. It differs from Kant's theory, as we have already noted, in that judgments of imputability are not directly related to questions about whether the action conforms to or violates duty, whether they are meritorious or blameworthy; that is, Hegel determines that some occurrence or fact is imputable to the agent prior to the question whether the agent is at fault or to blame for it, or deserves praise for it, or is the proper object of some other first-, second- or third-person attitude (e.g. first-person feelings of guilt, remorse, pride, or self-contentment, or second-person reproach or gratitude, or some third-person attitude of praise or condemnation).

This would allow Hegel to recognize, for example, the phenomenon that Williams calls “agent regret”—where agents feel badly, in a distinctively first-person way, about something they have done, or some result for which they are causally responsible, where no moral blame or fault attaches to them—Williams’ example is that of a truck driver whose truck has accidentally struck and killed a child, even though the driver was in no way at fault (not the least bit negligent or in any other way guilty or the proper object of any moral reproach) (Williams, 1981, pp. 26–33). In other words, Hegel’s theory acknowledges that moral imputation—what we call “responsibility”—extends farther than fault or blame (or their opposites). It concerns not merely praise- or blameworthiness, but rather the meaning of *who we are*, in the broadest sense, as moral agents, and it recognizes that this may involve our relation to occurrences that are a matter of luck and, more generally, lie beyond the scope of our free will.

10.8 Taking Responsibility

When we act in the world, we pursue various aims under conditions we did not determine, and the outcome often depends on factors we cannot control. If I am a rational agent, I know all this before I act. Hegel takes this knowledge, and the responsibility it involves, to belong to the very concept of moral subjectivity. As a rational agent, I accept my finitude, and my exposure to contingency, as a condition of acting. In so doing, I am *taking responsibility* for the results of what I do, even when these results are in some measure beyond my power. In choosing to act in a world of contingency, I also *choose* to expose myself to good or ill fortune, and I freely posit both good and ill fortune as the existence of my free will, giving them a right over me, making myself responsible for what comes about as their result. My actions may meet with good or bad luck, but in my very choice to act I opened myself up to both possibilities, and I *chose* to incur the blame if things turn out badly, just as I stand to deserve the credit if they turn out well. This exposure to fortune is simply a general condition of our action, and so we freely posit it along with any action.

An action is a *venture*: it may turn out well or badly, and by taking it we both seek a good outcome and take responsibility for a bad outcome, knowing from the start that it depends on factors beyond our control which way things will turn out. A mature and rational human being knows all this, accepts it, and therefore does not attempt to escape responsibility for bad results by drawing a distinction between good and ill fortune—claiming, for instance, that I should get as much credit when things have gone wrong as I would if they had gone right, or suggesting

(enviously) that a person who is more successful in doing good deserves no more credit than I do because his success was due to better luck.

Not all actions equally expose themselves to good or bad fortune. An action that merely complies with a strict duty makes a minimal venture. It exposes itself to the risk that what I do to meet my obligation may misfire, so that I fall short of meeting it. But beyond that, in doing what is required of me I assume very little responsibility for what may happen as a result, and I should get correspondingly little credit for surprisingly good results and bear little burden of responsibility for disastrously bad ones. On the other hand, if my action itself is more venturesome, something not strictly required of me that seeks to achieve some good I am not strictly required to produce, or that seeks to avert some harm that it is not already my duty to prevent, then (to put it in Hegel's terms) I "make good or ill fortune into the existence of my own free will" to a much greater extent.

Hegel's theory thus implies something vaguely analogous to the Kantian theory of imputability we examined at the beginning. Actions that comply with strict duties posit less of my freedom in good or ill fortune than acts that it is morally up to me whether to perform. I am less responsible for their consequences, less to blame if these turn out badly, and less to be credited if they turn out surprisingly well. For I was only doing what I had to do, hence not venturing as much. Projects that seek some good, analogous to Kant's meritorious acts, leave themselves more open to good or ill fortune; I get more credit if they turn out well, but must bear the responsibility if they turn out badly. Actions that are wrong are also ventures—but ventures in evil, and through them I open myself to blame for whatever harm I cause, even if (like Hegel's arsonist), I do harm that far exceeds what I intended.

Some philosophers, and even more often some criminally minded politicians, like to entertain the thought that a wrongful action might be justified, or at least in some way redeemed, if its consequences are good. Some of Hegel's remarks about the deeds of "world historical individuals" might even be seen as supporting this thought. But I do not think it would be correct to try to enlist Hegel in support of this thoroughly evil thought. He is quite stubborn in resisting every attempt to provide some kind of moral justification or excuse for any form of wrongdoing. In discussing the crimes of world-historical individuals spirit has made the means of attaining a higher ethical order (VG 171/141), Hegel is not referring to moral (or ethical) justification or responsibility at all. He is employing an "absolute" or "world-historical" species of justification that falls entirely outside both morality and ethical life (PR § 345). This is also not a justification any individual (e.g. Caesar) could ever give for his own actions, but belongs solely to the reflections of the philosophical historian, who is trying to grasp spirit's striving to realize its own nature in the course of human history.

Hegel's view is that from the standpoint of abstract right, morality, and ethics, world-historical individuals are (as the so-called "valet moralist" rightly says) merely criminals and villains; Hegel thinks they pay the penalty for their deeds and suffer the loss—they die early, like Alexander, are murdered like Caesar, or sent to St Helena like Napoleon (VG 105/89). Hegel never suggests that they have been treated worse than they deserve. If (like the valet moralist) we fail to see the supra-moral justification for their crimes, we do no moral injustice to Alexander, or Caesar, or Bonaparte; we merely deprive ourselves of the philosophical understanding of historical events we might have had, and fail to appreciate the rationally theodicy of history into which Hegel is fitting them. These justifications have nothing to do with the moral imputability of actions or consequences. As we have observed, Hegel lists the slogan "The end justifies the means" as fairly far along in the list of attitudes and ways of thinking that embody evil—it is worse than bad conscience, hypocrisy, and probabilism, and just above Fries' "ethics of conviction" (the corrupt theory that tries to pronounce justified an action that is objectively wrong if the agent performed it in the mistaken belief that it is right) (PR § 140R).

Hegel would be as reluctant as Kant is to impute merit to someone who had done wrong, but whose wrongful action, through the ironies of fortune, had produced good consequences. Rather than this, Hegel would sooner revisit the question whether the action had really been wrong after all, since an action whose outer results are "substantial" gives expression to an inner volition that is equally so: "The truth of the intention is only the deed itself." But such a revision, as we saw earlier, would be based on the *nature* of the action—a systematic consideration of the action and its results, and on the systematic results of actions of that kind, and not on the transitory, accidental consequences of a particular action on a single occasion.

Whether an action is right or wrong may sometimes depend on the results that may be imputed to the agent pertaining to actions of that nature. This is why Hegel cannot treat judgments about whether actions are right or wrong, dutiful or undutiful, as prior and independent conditions of the imputability of their consequences, as is done by Kant, and by the doctrine of double effect. But Hegel can make judgments of imputability in particular cases in part dependent on their rightness or wrongness, as when an action's wrongness belongs to its nature and therefore falls within the agent's purpose and intention, so that (as in the case of the arsonist) the agent bears the blame for all the bad consequences that belong (even if only contingently) to the nature of the action. And Hegel's theory advocates our acceptance of the moral ambiguity of actions that may be morally justified, but where the agent must also bear the burden of responsibility for doing foreseeable harm or intentionally bringing about bad consequences.

11

Marx on Equality

11.1 Inequality Today

We live in a world in which wealth, power, resources, life-prospects are very unevenly distributed. American society today is by many measures even more unequal than the capitalist order against which Karl Marx wrote in revolutionary protest in the mid-nineteenth century. Since Ronald Reagan's election in 1980, the wealthiest one percent of Americans have seen their incomes increase by 275 percent, while after accounting for inflation, the typical hourly wage for a worker has increased just \$1.23.¹ As of 2004, this privileged one percent owned more than five times as much as the total owned by the bottom half of the wealth distribution, and this inequality has continued to grow: in 2009–2010, as we began to recover from the banker-made financial disaster from which the bankers made billions, the top 1% of US income earners captured 93% of the income growth.² The United States is the most unequal of all developed nations, but the inequality between it and poorer countries is even greater. The trend to increasing inequality began in the 1970s; it is still accelerating, with no end in sight.

The maldistribution could reflect no conceivable measure of desert or distributive justice. It arises not from a “free” market (whatever that might be), but from a market characterized by monopolies and oligopolies, collusion, bought political influence, and cronyism, some of it criminal even under our extremely lax laws and regulations. This system of inequality is not merely self-perpetuating but self-reinforcing. Political institutions are still largely in the hands of those who benefit most from the inequality, and they are constantly being used to

¹ <http://www.nationofchange.org/bill-moyers-essay-crony-capitalist-blowout-1358002477>

² This was an estimate prepared in 2004 by Arthur B. Kennickell on behalf of the Federal Reserve Board: <http://www.wealthandwant.com/issues/wealth/50-40-5-4-1.htm>. As such inequalities have been growing steadily since then, it is a very conservative estimate of how things are now. For further statistics on inequality, see <http://inequality.org>.

widen, intensify, and multiply the disparities between the privileged and the disadvantaged.

It is understandable, therefore, that some among that tiny minority of us malcontented and politically impotent intellectuals who are outraged by this situation should have been busy exploring the idea that social, political, and economic equality is something greatly to be valued, either for its own sake or because it is indispensable for the attainment of other social goods, which our society so conspicuously lacks. A little over twenty years ago, G. A. Cohen expressed the egalitarian intuition this way: “I take for granted that there is something justice requires people to have equal amounts of, not no matter what, but to whatever extent is allowed by values which compete with distributive equality” (Cohen, 1989, p. 906; cf. Arneson, 1989; Temkin, 1993; Dworkin, 2000). Many others would not put it that way, but nevertheless consider themselves egalitarians, holding that equality should be valued either for its own sake or as a way of promoting other important goods, such as community, political democracy, and personal liberty and self-respect (see Miller, 1982, 1997; Scanlon, 1986; Sen, 1992; O’Neill, 2008; Hausman and Waldren, 2011).

Karl Marx opposed the systematic inequalities in the society around him. He also explicitly advocated, at least for the near future, many social measures that egalitarians also support: for example, a graduated progressive income tax, abolition of the right of inheritance, equal liability of all to labor, and universal free public education (CW 6:505). So it is natural for us to think of Marx as an egalitarian of some sort. But Marx definitely did not share the egalitarian intuitions I have just been describing. My task here is to try to understand why he did not.

11.2 Marx and Engels on the Meaning of Equality

We should begin by noting a subtle difference between the writings of Marx and those of Engels whenever the concepts of equality and inequality come up. Marx, quite frequently, and with very few exceptions, mentions “equality” only to make the point that it is an exclusively *political* notion, and, as a political value, that it is a distinctively *bourgeois* value (often associated with the French revolutionary slogan “*liberté, égalité, fraternité*”). Far from being a value that can be used to thwart class oppression, Marx thinks the idea of equality is actually a vehicle for bourgeois class oppression, and something quite distinct from the communist goal of the abolition of classes (CW 3:79, 163–4, 312–13, 4:39–41; 5:60; 6:228, 511; *Capital* 1:280). Engels, on the other hand, while also repeating these claims, is sometimes more positive in his attitude towards equality. Then he distinguishes “political equality” or “equality of rights” from “social equality” or “real equality” (CW 3:393–4; 6:5–7,

19, 28–9, 6:346; 10:414; 24:286–7; 25:19–20, 592). There are a few similar passages, though only a few, that are either by Marx or co-authored (CW 3:79, 163; 5:479). But the differences between Marx and Engels can be seen to narrow when Engels explains further the proletarian demand for “social” or “real” equality:

The demand for equality in the mouth of the proletariat has therefore a double meaning. It is either—as was the case especially at the very start, for example in the Peasant War—the spontaneous reaction against the crying social inequalities, against the contrast between rich and poor, the feudal lords and their serfs, the surfeiters and the starving; as such it is simply an expression of the revolutionary instinct, and finds its justification in that, and in that only. Or, on the other hand, this demand has arisen as a reaction against the bourgeois demand for equality, drawing more or less correct and more far-reaching demands from this bourgeois demand, and serving as an agitational means in order to stir up workers against the capitalists with the aid of the capitalists’ own assertions; and in this case it stands or falls with bourgeois equality itself. In both cases the real content of the proletarian demand for equality is the demand for the *abolition of classes*. Any demand for equality which goes beyond that necessarily passes into absurdity (CW 25:99).

Here Engels regards the proletarian demand for equality as standing or falling with the bourgeois demand, and, when it goes beyond the demands of bourgeois equality, as drawing conclusions of doubtful validity. The real meaning of this demand, he thinks, to the extent that this demand has any validity, is the demand for the abolition of classes. In fact, Engels regards the proletarian demand for equality as valid only at a stage of development which he regards as now past:

The idea of socialist society as the realm of *equality* is a one-sided French idea resting upon the old ‘liberty, equality, fraternity’—an idea which was justified as *stage of development* in its own time and place but which, like all the one-sided ideas of the earlier socialist schools, should now be overcome, for it produces only confusion in people’s heads and more precise modes of presentation of the matter have been found (CW 24:73).

Marx did include in the *Rules of the International Workingmen’s Association*, as among aims of the Association, the following words: “...equality of rights and duties and the abolition of class rule” (CW 20:14; 23:3). It seems clear that he accepted “equality of rights and duties” because others insisted on it, and he thought it did no harm if followed immediately by “the abolition of class rule,” which he took to be a better expression of the same aim.

Marx and Engels distinguish “abolition of classes” from “equalization of classes,” regarding the latter as a recipe for “harmony between Capital and Labor” (of course, on Capital’s terms, since those terms define the production and class relation). They reject “equalization of classes” on that ground (CW 23:88).

In treating the notion of equality, then, Engels as well as Marx holds fundamentally to two ideas: *first*, that equality is properly speaking only a political notion,

and even a specifically bourgeois political notion; and *second*, that the real meaning of the proletarian demand for equality, to the extent that it has a meaning, is the demand for the abolition of classes—and that this demand is a better developed and more precise expression of proletarian aspirations. To understand Marx's reasons for rejecting the common intuition that social justice, in some desirable sense of the term, requires equality, we need to explore further these two ideas.

11.3 Equality as a Political Concept

The Marxian idea that equality is a political notion is itself a complex idea—as complex as Marx's understanding of the political itself. The most basic bourgeois equality, as Marx understands it, is a form of “procedural equality,” namely, *equality before the law*: the legal system must not accord some estates more privileges than others (as was still true in the feudal-aristocratic political orders of early modern Europe) (CW 6:228; 24:286).

As John Rawls long ago pointed out, mere procedural equality—“treating like cases alike”—is very weak in what it demands. Depending on the grounds that may be offered to justify inequalities, it is consistent with caste systems or even slavery (Rawls, 1971, pp. 507–8). Equality before the law, excluding privileges based on caste or estate, is stronger. But Marx clearly thinks that such procedural conceptions of equality tailored to bourgeois legal and political institutions will merely support the class oppression to whose divisions these bourgeois institutions correspond. Thus we find in Kant, for instance, the identification of equality with “independence of being bound by others to more than one can in turn bind them” (MS 6:237; cf. 6:314; TP 8:291–4, 297; EF 8:349–50). This equality, Kant says, “is quite consistent with the greatest inequality in terms of the quantity and degree of their possessions” (TP 8:291–2). Kant's own arguments that the poor have a right that the wealthy should be taxed to support them are based not on considerations of equality, but rather of their right to freedom and independence as their own master (*sui iuris*) (MS 6:325–6; TP 8:295). Fichte's even more far-reaching demands for economic redistribution are based not on equality but on the right to be able to live independently from one's own property (NR 3:212–15; GH 3:402–3). Both Kant and Fichte here are following Rousseau, who argues that equality is necessary only because freedom cannot exist without it:

If one inquires into precisely what the greatest good of all consists in, which ought to be the end of every system of legislation, one will find that it comes down to these two principal objects: freedom and equality. Freedom, because any individual dependence is that much force taken away from the state; equality, because freedom cannot subsist without it. . . . As for wealth, no citizen should be so very rich that he can buy another, and none so poor that he is compelled to sell himself (Rousseau, 1997, II, 11 [1, 2], p. 78).

The wisest philosophers in this tradition thus do not regard inequality in possessions, benefits, or opportunities as bad in itself, but do think it necessary to limit it for other ends, especially freedom (independence of the arbitrary will of another). Marx disagrees with this bourgeois tradition on many points, but fundamentally agrees with it on this one.

Marx's view that equality is a bourgeois notion, involving only equality before the law and formal equality in contractual dealings is what underlies, I believe, Marx's own theoretical requirement in *Capital* that surplus value must be explained on the assumption that equal values are exchanged between formally free and equal economic agents (see *Capital* 1:271, 301). Marx's claim that the sphere in which the capitalist purchases labor power is "a veritable Eden of the innate rights of man . . . liberty, equality, property and Bentham" (*Capital* 1:280), is quite literally meant, however ironical its intent.

More generally, it ought to be hard to miss the fact—and also impossible to interpret it away—that Marx does not regard capitalist exploitation of labor as unjust, or as any violation of the laborer's rights (see Wood, 2004, Chapters 9–10, 16). In Marx's view, the only rights that could come into question here are those corresponding to the bourgeois mode of production (*Capital* 1:301; cf. *Capital* 3:460–1). Right (*Recht, droit*), for Marx, as for Rousseau, Kant, and Fichte, is a concept essentially associated with political and legal institutions, which, on Marx's historical materialist theory, are merely the legal-political superstructure that arises out of its real foundation in the existing mode of production (CW 29:263). To attempt to apply under the conditions of bourgeois society any standards of right but those corresponding to the capitalist mode of production is, in the words of Engels quoted above, to make "a demand that necessarily passes into absurdity."

We can see such views in action in Marx's critique of the Gotha Program's demands for "a just distribution," and "a distribution of the proceeds of labor to all members of society with equal right." On the Program's demand for a "just distribution," Marx comments with a series of pointed rhetorical questions:

Do not the bourgeois assert that the present distribution is "just" [*gerecht*]? And is it not in fact the only "just" distribution on the basis of the present day mode of production? Are economic relations regulated by concepts of right [*Rechtsbegriffe*], or do not, on the contrary, relations of right arise out of economic ones? (CW 24:85–6).

Marx takes the answers to these questions to be plain: of course the bourgeois *do* assert that the present distribution is just—and Marx agrees with them that it *is* the only just distribution on the basis of the present-day mode of production. He agrees, because the materialist conception of history says that economic relations are *not* regulated by concepts of right, but, on the contrary, relations of right *do* arise out of economic ones. Marx then continues:

Right can never be higher than the economic structure of society and its cultural development conditioned thereby... Any distribution whatever of the means of consumption is only a consequence of the distribution of the conditions of production themselves. The latter distribution, however, is a feature of the mode of production itself... If the elements of production are [distributed as they are under the capitalist mode of production], then the present day distribution of the means of consumption results automatically (CW 24:87–8).

This means that the only standards of distribution that can apply in capitalist society are those that result in the capitalist distribution of wealth. As Marx puts it in *Value, Price and Profit*: “To clamor for *equal or even equitable remuneration* on the basis of the wages system is the same as to clamor for *freedom* on the basis of the slavery system” (CW 20:129).

11.4 The Defects in Any Equal Standard

As we have seen, Cohen takes it to be self-evident that justice requires something to be distributed in equal amounts (setting aside values that might compete with equality), but egalitarians involve themselves in squabbles and perplexities as soon as they ask what the equal standard should be. Some think it should be welfare, others wealth or income, others opportunity or capabilities. Other philosophers recognize that there are multiple dimensions on which equality might be measured. Inequalities, especially on the side of disadvantage, tend to form clusters: low income often goes along with poor education, poor health, lack of control over one's circumstances, lack of political influence, and so on. There are obviously causal interconnections here too, but how do they work? Sometimes, however, these bad circumstances also come apart, and if we equalize along one dimension, we may fear we are unfairly neglecting those who are disadvantaged along other dimensions. These considerations lead people to ask: “Which of these forms of disadvantage should be included in our measures of inequality, and how much weight should be given to each, in deciding what society should try to equalize for its members?” Then still other philosophers worry about equalizing in matters for which individuals ought to take responsibility for the actions that make them better or worse off than others. It might be good to equalize, they argue, but not in *these ways*.

Do we really have any idea at all what it is that justice requires that it should be equal for everyone? Even if we doubt that we do, the view still seems to be commonly held that if we could only find the right equal standard, equality of something is still the demand of justice—or at least one of its demands. But if for some *x*, whatever *x* may be, justice truly demands that everyone have an equal amount of *x*, then it looks like an *injustice* already to let any other value permit us to distribute

more of x to some than to others. So Cohen's original intuition itself strikes me as potentially self-contradictory.

Marx's response to all these difficulties is simply to reject the egalitarian intuition, and deny there is any equal standard that could be used to formulate some ideal demand of justice. At the same time, Marx recognizes that some equal standard is likely to be applied in fact as long as the (still essentially bourgeois) notions of right and equality are in play. He accepts that this will have some consequences that can be regarded only as unsatisfactory. (Justice is not for Marx, as it is for Rawls, the first virtue of social institutions; it is simply a feature they display when viewed from a legal and political standpoint.)

The *Critique of the Gotha Program* also deals with the system of distribution Marx expects to prevail under the first phase of post-capitalist society: "a co-operative society based on common ownership of the means of production."

The individual producer receives back from society—after the deductions have been made ["for replacement means of production, expansion of production, reserve or insurance funds to provide for accidents and funds for those unable to work"] exactly what he gives to it. What he has given to it is his individual quantum of labor... He receives a certificate from society that he has furnished such and such an amount of labor... and with this certificate he draws from the social stock of means of consumption as much as costs the same amount of labor... [Equal right] is therefore a right of inequality in its content, like every right (CW 24:86).

Marx then emphasizes that this is still "equal right" only in the bourgeois sense of the term, even if the standard is no longer bourgeois. It remains unsatisfactory *precisely because* it applies an equal standard.

Right by its very nature can consist only in the application of an equal standard; but unequal individuals (and they would not be different individuals if they were not unequal) are measurable by an equal standard only insofar as they are brought under an equal point of view, are taken from one definite side only, for instance, in the present case, are regarded *only as workers* and nothing else is seen in them, everything else being ignored. Further, one worker is married, another is not; one has more children than another, and so on and so forth. Thus with an equal performance of labor, and hence an equal share in the social consumption fund, one will receive more than another. To avoid all these defects, right instead of being equal would have to be unequal. But these defects are inevitable in the first phase of communist society (CW 24:86).

To avoid the defects that are inevitable in any system of equal justice, Marx thinks, right instead of being equal, would have to be unequal. And he looks forward to a more distant future in which he hopes people will not have to think in terms of right or justice at all:

In a higher phase of communist society, after the enslaving subordination of the individual to the division of labor, and therewith also the antithesis between mental and physical labor

has vanished; after labor has become not only a means of life but life's prime need; after the productive forces have also increased with the all-round development of the individual, and all the springs of co-operative wealth flow more abundantly—only then can the narrow horizon of bourgeois right be crossed in its entirety and society inscribe on its banners: *From each according to his abilities, to each according to his needs!* (CW 24:86–7).

This last slogan is now popularly associated with Marx himself, but at the time, and for Marx's intended audience, it would have been associated with Louis Blanc (Blanc, 2012). The more remote source of the slogan, however, may also be the New Testament: "And all that believed were together, and had all things common; and sold their possessions and goods, and parted them to all men, as every man had need" (Acts 2:44–5). The slogan might therefore just as well be attributed to St Luke as to Louis Blanc or Karl Marx—and also held up to Bible-believing Christians as a principle to which their faith commits them.

Be that as it may, Marx uses it here precisely because it is *not* a principle of equal right in any sense: neither people's abilities nor their needs are equal: both differ qualitatively, in ways that may not be measurable (or, therefore, equalizable); and even to the extent that they do admit of measurement, they may differ greatly. People's abilities may vary greatly as to the amount of their contribution to the needs of society, and people's needs may differ greatly according to the amount of social resources required to meet them. A society that lived according to Louis Blanc's slogan would not be applying an equal standard either in what it asks of individuals or in what it distributes to them.

Perhaps Louis Blanc's slogan seems egalitarian to us because we are tempted to think it requires *equally* of people that they contribute their abilities, and it distributes to people *equally* what they need. We might have the same thought about the *Communist Manifesto*'s proclamation that socialist society would be "an association in which the free development of each is a condition of the free development of all" (CW 6:519). For it too provides *equally* for the free development of each individual, even though the conditions for the free development of different people might, like their abilities and their needs, differ greatly, both qualitatively and quantitatively, from one individual to another. But our thought that this involves any genuine equality rests on an elementary confusion. All that is true here is that the reference to "each" implies that there is some predicate that will apply *equally* (that is, universally or without distinction), to all the individuals falling under the pronoun "each." The use of "equal" or "equally" in this context implies no substantive egalitarian principle whatever; for the predicate in question may be one that involves a highly unequal treatment of these individuals. Thus one might say of a caste system, in which some castes are privileged and others severely disadvantaged, that it treats everyone *equally* in assigning to them what pertains to

their caste position. As I once put this point: “One might as well say that the Grim Reaper is an egalitarian on the ground that once we die, we are all *equally* dead” (Roemer, 1986, p. 296).

Of course, neither Louis Blanc’s slogan nor the *Manifesto*’s separate individuals into castes, but both slogans do entail unequal treatment of individuals, to the extent that their abilities, needs, and conditions of free development might be qualitatively different and quantitatively unequal. Nor is Louis Blanc’s slogan for Marx a matter of right, for it applies only “after the narrow horizon of bourgeois right [has been] crossed in its entirety.” And if right is a matter of state coercion, it would be extremely objectionable (and clearly not something intended by either Louis Blanc or Karl Marx) that the state should attempt to extract coercively from each person every last bit of social contribution that their abilities might afford.

11.5 Equal Right and the Political State

This last point enables us to see that egalitarianism, regarded as equal right, is usually a political notion in another way that hasn’t yet been made fully explicit. When people say that justice requires treating people equally from some point of view, or that people have a right to equal amounts of something, the assumption (whether tacit or explicit) is that this treatment or this doling out are going to involve actions of the political state. The basic notion of bourgeois equality is equality before the law. In bourgeois society, this means equality in political terms that correspond to the laws of the bourgeois economy. As Engels puts it: “Political equality—what is it but the declaration that class differences do not concern the state, that the bourgeois have as much right to be bourgeois as the workers to be proletarian?” (CW 23:418–19).

It is usually taken for granted that economic redistribution, through tax policy, education policy, land reform, and other measures will be carried out through laws and their administration. In whatever way people are being treated as equals, they are going to be regarded as equals fundamentally as citizens of some political entity, such as a nation state—only by some extension of this can one speak of equality in international terms. Marx’s deepest reasons for questioning the notions such as right, justice, and equality, is that these notions apply to people only in their specifically political identity. But ever since his early essay *On the Jewish Question*, Marx had the deepest reservations about that way of considering human beings. (For a good discussion of the theme of the political in this essay, see Wolff, 2002, pp. 40–7; cf. Wood, 2004, pp. 51–8, 63–6).

Marx’s teacher Bruno Bauer argued that Jews should not seek political emancipation until they first emancipate themselves from their religion (as indeed

Christians too must do if they are genuinely to be emancipated). In his review essay, Marx objects that Bauer has not understood the nature of political emancipation itself, and its relation to human emancipation (CW 3:149). "We do not say to the Jews as Bauer does: you cannot be emancipated politically without emancipating yourselves radically from Judaism. On the contrary, we tell them: because you can be emancipated politically without renouncing Judaism, political emancipation is not human emancipation" (CW 3:160).

For Marx, human emancipation must belong to that sphere which Hegel called "civil (or bourgeois) society" (*bürgerliche Gesellschaft*), which is where human beings lead their truly human life.

The political state [says Marx] stands in the same opposition to civil society, and prevails over the latter in the same way as religion prevails over the narrowness of the secular world. . . . [Here] the human being. . . leads a twofold life, a heavenly and an earthly life: life in the *political community*, in which he considers himself a *communal being*, and in *civil society*, in which he acts as a private individual, regards other men as means, degrades himself into a means, becomes the plaything of alien powers. . . . In the state, where the human being is regarded as a species-being, he is the imaginary member of an illusory sovereignty, is deprived of his real individual life and endowed with an unreal universality (CW 3:154).

When people think of their communal life in terms of rights and justice, they are thinking of their social nature as their political nature, which relegates their real social life in civil society to the status of a private, atomistic, and merely self-interested life—the sort of life it actually assumes in bourgeois capitalist society. Their rights as members of the state, therefore, are seen by them not as a positive social or species-life, but only as a set of powers they have over against other human beings—the power to free themselves from the control of others, or to obtain from others, via state coercion, what they can claim by right.

None of the so-called rights of man, therefore, go beyond egoistic man, man as a member of civil society, that is, the individual withdrawn into himself, into the confines of his private interests and private caprice, and separated from the community. In the rights of man he is far from being conceived as a species-being; on the contrary, species-life itself, society, appears as a framework external to individuals, a restriction on their original independence. The sole bond holding them together is natural necessity, need and private interest, the preservation of their property and their egoistic selves (CW 3:164).

The political life of human beings, therefore, is not their real social life, and political emancipation is not human emancipation. Equality, along with right, justice, and other conceptions of a merely political nature, are necessarily inadequate expressions of the human aspiration to membership in a free community. This is why these conceptions are also inadequate to genuine human emancipation. This is the point, therefore, at which we should turn to the second of the two Marxian

ideas mentioned above—that the true meaning of “equality” when used by the proletariat is the demand for the *abolition of classes*.

11.6 Class Society

It is probably natural for us to think of the Marxian notion of a classless society in egalitarian terms. Accordingly, we naturally interpret the Marxian notion of class oppression or exploitation as a particularly odious form of inequality, and we see the remedy for it (whatever name Marx may choose to give it) as egalitarian in substance and content. The rest of this and the next section will be devoted to explaining why these natural thoughts are profoundly mistaken, at least as an interpretation of Marx.

Marx does not tell us very much about what a classless society would be like. He scorns the enterprise of “writing recipes for the cook-shops of the future” (*Capital* 1:99) and insists that communism “is not an ideal to which reality has to adjust itself. We call communism the real movement which abolishes the present state of things” (CW 5:49). Marx takes this position because, to a perhaps surprising extent, he accepts Hegel’s view that “the owl of Minerva begins its flight only at dusk” (Hegel, PR Preface, p. 23)—in other words, that future human history is necessarily largely opaque to us, though Marx thinks we can perceive those tendencies in the present toward future developments, and choose to align ourselves with those we favor. We learn little in Marx about the socialist, communist or classless society of the future, except through what we learn from him about the class society of the present day, and those aspects of human life he thinks depend on social classes and class antagonisms.

Marx began a chapter on classes in Volume 3 of *Capital*, but the draft breaks off after only a couple of paragraphs. The only point he makes is the negative one that classes are *not* to be distinguished merely on the basis of different sources of revenue (e.g. landowner: rent; capitalist: profit; worker: wages). “From this standpoint, physicians and officials, for instance, would also form two classes... The same would hold for the infinite fragmentation of interests and positions into which the division of social labor splits laborers and capitalists” (*Capital* 3:1025–6). Elsewhere, however, Marx tells us more about classes. They arise out of production relations, he says, because these relations “create masses with a common situation, common interests” (CW 6:211). But such masses, with opposing interests, become classes, in Marx’s view, only when they engage in some kind of common action sufficient to make them potent forces in history—either in the role of ruling class or revolutionary class, or at least as a subordinate player in the struggles between the main classes that determine a historical epoch or the revolutionary transition between epochs. The common situation of a class, that give its members shared

interests, makes it into a class potentially or in itself; it becomes a class actually or for itself, or “constitutes itself as a class,” only when it acts as a class, through common social, political, or intellectual deeds (CW 6:211; CW 11:187). “Only then do the interests it defends become class interests” (CW 6:211). Class interests, moreover, always consist in an opposition to other class interests. “Separate individuals form a class only insofar as they have to carry on a common battle against another class” (CW 5:77). The very existence of classes is constituted by a class struggle (CW 6:482). A class society, in its very concept, is one in which the interests of some are irreconcilably opposed to the interests of others. In such a society, any conception of general interests, or of universal values or principles having normative authority for all members of society, is necessarily an illusion: typically, it is the illusion of the ruling class. This is why “law, morality, religion, are [to the clear-sighted proletarian, only] so many bourgeois prejudices, behind which lurk in ambush just as many bourgeois interests” (CW 6:494–5).

The process by which a class arises out of a set of production relations is a process necessarily involving people’s ideas and consciousness, and also a process involving activities (especially political ones) that pertain to the social superstructure, rather than to the economic foundation of society. For this reason, it is a fundamental misunderstanding to think that Marx regards only economics as determining human history, and treats politics, law, morality, and religion as merely “epiphenomenal.” The language of “superstructure” and the Marxian phrase “social being determines consciousness” are not about the effectiveness or ineffectiveness of that to which these terms apply, but instead about their opacity to those individuals that act in terms of them. The point is that superstructural activities are not typically understood for what they are by those subject to them: class-consciousness does not typically employ the concept of “class” at all. The most characteristic concept it employs, in fact, is precisely that of *general* or *universal* interests, principles, laws, norms, or values: for example, those of right or justice, morality, or the common interests of society on which the authority of a political state is thought to rest. This happens when the interests of a particular class “develop in spite of the persons into common interests, standing independently over against the individual persons and in this independence assuming the form of *general* interests” (CW 5:245).

On the different forms of property, on the social conditions of existence, here arises a whole superstructure of different and characteristic feelings, illusions, modes of thinking and views of life. The whole class creates and shapes them from its material foundations and out of the corresponding social relations. The single individual, to whom they flow through tradition and education, can imagine that they are the real motives and starting point for his action (CW 11:128).

And as far as his own psychological motives are concerned, he may be right in imagining this. It is not a matter of self-deception about his own motives, as if he were mistaking selfish class interests for the true values that he thinks are motivating him. The illusion rather consists in not understanding the true social source and power of these ways of thinking, ascribing to them an objective meaning and authority they don't have—because, in fact, nothing ever has it.

It is *not* Marx's view that bourgeois conceptions of justice, right, morality are merely class interests usurping the place of the true or genuine standards (which, we might think, Marx would identify with proletarian standards, or the standards that will apply in some future communist society after classes have been abolished). Marx holds, instead that the whole concept of general interests, universal principles, or values, anything that might claim authority over individuals in the name of some common interest or objective truth about right or morality, is never anything but the interests of some particular class falsely claiming such an authority. The very concept of universal moral authority, in Marx's view, is a product of class society and would have no application were it not for its mystifying and ideological use in class struggles. A class interest falsely claiming universal validity or authority is the closest thing there ever is, or ever could be, to anything actually having such validity or authority.

This is why, in Marx's view, individuals do sacrifice their individual egoistic interests to class interests (those of their own class, or even sometimes of a class hostile to them). And this is just as true of proletarian class interests as of any others. The difference is—and here Marx gives to capitalism a great deal of credit—that in modern bourgeois society “man is at last compelled to face with sober senses, his real conditions of life and his relations with his kind” (CW 6:487). It therefore becomes possible for communist consciousness to accept the fact that it struggles, and even makes individual sacrifices, not in the name of any universal interest or objective principle of right or justice, but simply on behalf of the class interests of a revolutionary class: “Communists know very well that under determinate relations egoism as well as self-sacrifice is a necessary form of the successful interaction of individuals . . . [Both] are sides of the personal development of individuals, equally generated by the empirical conditions of life” (CW 5:246–7). Nor does Marx think that egoism is in general any more or less “rational” than self-sacrifice on behalf of the interests of some class. Both motives are equally products of the social conditions of life, and once individuals come to understand this, it is up to them, as free individuals, to decide by which their actions should be motivated.

Marx thinks that social relations of the past—in the modern world, especially bourgeois social relations—have created the prejudice, or illusion, that egoism is more “rational” than self-sacrifice for a revolutionary cause, as well as the equally

mystified illusion that self-interest can be rationally overridden only by some universal or objective interest or value. Neither egoism nor class interest has more “natural” authority than the other—nor, for that matter, does either have any *natural* authority at all. Such talk belongs entirely to the historically conditioned illusions to which class society has made us susceptible. All such motivational patterns have only the rational authority that the clearsighted thinking of individuals might give them. What Marx is saying in the above passage is that communists, once they come to understand social life and history materialistically, will be freed from all such illusions. This is the liberation that becomes possible for us when we face with sober senses our true relations with our kind.

11.7 Marx and Stirner

There is nevertheless a temptation to seek in Marx (or read into him) some conception of a system of rightful distribution that will govern future communist society—what Engels once referred to as “a human morality which stands above class antagonisms” (CW 25:88). Marx, however, never uses such phrases, and to the imagined charge that “communism abolishes...all religion and morality, instead of constituting them on a new basis,” the *Manifesto* replies only that “the Communist revolution is the most radical rupture with traditional property relations; no wonder that its development involves the most radical rupture with traditional ideas” (CW 6:504). The response, in other words, is that the “abolition of all morality” is part of that radical rupture.

The unpublished manuscript of *The German Ideology*, which Marx and Engels willingly left (as Marx says) to the gnawing criticism of the mice (CW 29:263), devotes well over three hundred pages of often tedious polemic to attacking Max Stirner’s 1844 book *The Unique Individual and His Property* (Stirner, 1995). It is seldom appreciated how far they had gone in the earlier pages of *The German Ideology* (and how far Marx himself went in all his writings) towards accepting some of Stirner’s more radical ideas.

The standard English title of Stirner’s book is *The Ego and Its Own*. This title is no doubt meant to capture the fact that Stirner describes his position as “egoism.” But “Einzig” means “individual” or “unique,” and it is essential to Stirner’s view that the free individual is entirely unique, self-defined, not subject to any universal standards, whether of right, morality, or even general human self-actualization, flourishing, and well-being. Stirner’s book was a critique of everything that he thought served to enslave the individual personality, to subject it to any form of “hierarchy” or the “dominion of thoughts,” which deprive it of what is authentically its own. This critique begins with religion, but extends to all morality, even

including all ideals of human flourishing, and takes in all forms of social authority, family, community, state, or party, which might claim precedence over the egoism of the unique individual.

Marx clearly rejected some basic elements of Stirner's creed, such as his designation of "egoism" as the free individual's orientation and his rejection of all social ties that could not be seen by the egoist as merely forms of his own "self-enjoyment." But Marx adopts from Stirner the idea that free individuals are not subject to universal standards; in a classless society they will be merely the (social) individuals they are. Marx accepted Stirner's idea that all interests, ideals, and principles that claim universal authority are to be rejected as *ideology* in a sense equated with "the dominion of thoughts" and are therefore false impositions on human freedom (CW 5:24, 43–5, 59–61). This false universality is now interpreted by Marx and Engels as an expression of a society divided into warring classes; it is the way class interests try to impose themselves on us as having some sort of transcendent or sacred authority (CW 5:46–7, 61–3). A society that has transcended class antagonisms, therefore, would not be one in which some truly universal interest at last reigns, to which individual interests must be sacrificed. It would instead be a society in which individuals freely act as the truly human individuals they are. Marx's radical *communism* was, in this way, also radically *individualistic*.

Only within the community has each individual the means of cultivating his gifts in all directions; hence personal freedom becomes possible only within the community. In the previous substitutes for community, in the state, etc., personal freedom has existed only for the individuals who developed under the conditions of the ruling class, and only insofar as they were of this class (CW 5:78).

The motivations of free individuals in a genuine community might be described either as egoistic or altruistic, or rather, as Marx puts it in his excerpt-notes of 1844, they would be both at once, because for social individuals there is a natural harmony or even identity between what actualizes me, fulfills my needs, and what actualizes others or fulfills their needs, and at the same time actualizes the species-being that belongs simultaneously to myself and others: "In the individual expression of my life I would have directly created your expression of your life, and therefore in my individual activity I would have directly *confirmed* and *realized* my true nature, my human nature, my communal nature" (CW 3:228).

Communism differs from all previous movements in that it overturns the basis of all earlier relations of production and for the first time treats all naturally evolved premises as the creations of hitherto existing human beings, strips them of their natural character and subjugates them to the might of the united individuals (CW 5:81).

It is the association of individuals . . . which puts the conditions of the free development and movement of individuals under their control (CW 5:80).

There is no reason to think that Marx believed that the abolition of class society would do away with all sources of conflict between individuals, or bring them into total agreement on how to direct their collective future. Post-class society for him is not the end of history, but only the end of human “pre-history” (CW 29:264). But Marx thinks that in all past society (beyond the most primitive stages of economic development), these conflicts and disagreements have been determined in both form and content by the pervasive fact of class conflict. What people have represented to themselves as a “war of all against all” or “unsociable sociability” arising from human nature itself is, for Marx, as it was also for Rousseau, not a fact of nature but a social product. But human history has been even more deeply a history of human co-operation than of the antagonistic forms this co-operation has assumed. Marx holds that humanity has the chance to retain the social co-operation without class conflict, because “all previous historical movements were movements of minorities, or in the interest of minorities. The proletarian movement [however] is the self-conscious independent movement of the immense majority in the interest of the immense majority” (CW 6:495). This is why he thinks the proletariat has the opportunity to abolish class society and begin *human* history.

11.8 Coming to Terms with Marx

My aim so far has been to present *Marx's* views, and draw out their implications for what he thinks about the causes of social inequality and about ideals of social equality. I have even been trying to defend these views, at least in a limited and conditional way, by arguing that they make coherent sense, and by presenting Marx's reasons for holding them. But since I don't myself accept the entire Marxian story, I suppose I owe it to you (as well as to myself) to say something about where I myself stand in relation to all of this. Since I began seriously studying Marx in the mid-1960s, I have always found his basic critique of capitalism entirely convincing. Nothing that has happened in the past half-century has budged me a single iota from that conviction. But I have never been attracted only to that part of Marx I found compelling or credible. And some of the themes in Marx I have been most concerned to emphasize in my work on him do not involve ideas I have ever accepted myself. The following are the best explanations I am able to come up with for that curious fact.

Next to the sick, abominable, unthinking hostility toward Marx's ideas that prevails in much of the world, and especially in my hopelessly benighted country, the most contemptible obstacle to their sympathetic reception has always been the dogmatic pseudo-religious attitude of uncritical acceptance found among many

of his self-appointed followers, especially those that have been best organized and most resolute. I think I am drawn to the most radical and adventurous of Marx's teachings—those a sensible person can be least comfortable adopting uncritically—precisely because I want to distance myself from this second way Marx's thought has been abused, just as much as from the first way.

Then too, Marx's critical attitudes toward right and justice, and his radical rejection of all universal moral standards, have always had, if not a direct appeal, at least a special kind of attraction for me. I am attracted to these radical ideas because, even as I find myself unable to swallow them whole, they do seem to me to make a kind of sense. They certainly involve a form of metaethical antirealism—if not about all values, then at least about those values associated with right and morality. I find it refreshing that Marx frankly and openly accepts the radical rejection of all morality that plainly and necessarily goes with any such metathetical position. Here, very much to his advantage, Marx stands in sharp contrast to the squalid dishonesty found among ethical “emotivists,” “projectivists,” “fictionalists,” “quasi-realists,” and others who embrace essentially the same view, but then proceed to quibble and prevaricate in cowardly evasion of the radical conclusions that obviously do follow from their bleak moral nihilism.

As far back as I can remember, I have always thought, contrary to Marx, that there are objective standards of right and ethics, which are not mere masquerades worn by class interests (or by any other sort of subjective conation, feeling, or attitude). There is an unacknowledged tension in Marx's own views at this point, since he holds that some standard of right (distributive justice) must hold in the early phases of socialist society, but offers no positive account of their normativity—except to say that they will correspond to the then prevailing mode of production. And he even points out their defects in those standards themselves. Perhaps this is no different from the accounts of similar non-consequentialist moral norms that are given by some consequentialists, or the accounts of moral norms generally that are given by “expressivists” and “quasi-realists.” Those who offer such deflationary theories seem to think there is no incoherence in proposing to defend a moral standard in a way that would directly undermine it. The difference, which puts Marx's views in tension with themselves, is that Marx was honest enough to draw deflationary conclusions from deflationary arguments.

Be that as it may, for almost as long as I've believed there are objective standards of right, I have thought that Marx's correct account of capitalist relations and capitalist society can be used to show that the capitalist exploitation of labor violates the rights of workers—not on any grounds of *equality*, but because workers have

a right to lead their lives free from the coercive power that capitalists have always exercised over them in really existing capitalism. I accept the idea that human beings have dignity, and that this makes them equal *as persons*. This does not imply that they must be treated alike, or that there is anything of which all people must be given equal shares.

It is a significant development, however, to accept the principle that all human beings should address one another on equal terms, that there should be no distinctions of rank or self-worth among them. This belongs centrally to the Classical German tradition, through Kant's principle that humanity in persons is an end in itself, and the theory of recognition originated by Fichte and developed by Hegel. Marx, I believe, tended to take for granted this development, as one of the achievements of modern bourgeois society. He did not fully appreciate that even after bourgeois equal recognition is established in principle, there are still very significant rear-guard actions still to be fought in modern society based on this kind of equality, surrounding issues of race, ethnicity, and gender, some of them even at odds with the concrete conclusions drawn (or rationalized) by those philosophers who best established the principle that human beings are all of fundamental equal worth. I call these "rear-guard actions," but many of them are difficult, complex, and possibly even interminable. (There are, for example, now more African-American males incarcerated in U.S. prisons than were enslaved in the mid-nineteenth century.) Marx was also well aware that bourgeois class oppression can also easily create new forms of deficient recognition, but he still regarded the assertion of human equality in working-class programs as a bourgeois principle, something more to be tolerated than emphasized.

Nevertheless, it is important to distinguish equality of recognition—equality of self-worth among all human beings—from the idea that people should all have equal claims to something or that there is something of which everyone ought to be given equal share. Louis Blanc's slogan appeals to me, as it did to Marx, in large part because of its directly *anti*-egalitarian implication that in the end, society should *not* treat people equally. It is self-evident that any decent society should demand more of those who have more to give, and provide more to those whose needs are greater. It is not the least of the evils of capitalism that it has corrupted people, making this truth seem less self-evident to many of them than it is. Those to whom it does not shine as brightly as the sun ought to have to go and live in the wilderness, deprived of all the advantages of social life, until their vision clears. Further, in order to protect the freedom that human dignity requires, it seems equally evident that a decent society should be aggressive in protecting the freedom of the

vulnerable, which demands sharp restriction of the freedom of those in a position to take advantage of them.³

But to return, at the end, to my settling accounts with Marx: it has taken me a long time to realize that my biggest disagreement with him is over capitalism, of which his opinion was *far too favorable*. Marx thought capitalism was a transitional economic form, whose historic mission was to elevate the productive powers of humanity—albeit at terrible human cost—to the point where they would offer abundance to all in a higher and freer society. Marx saw that the capitalist system was subject to periodic instabilities and crises, from which he inferred that it could be only a transitory form of social organization. He also saw that it was extremely wasteful of the resources it put at people's disposal, arbitrarily giving great advantages to a privileged few, while condemning the masses whose labor produces social wealth to lives that are both materially and spiritually impoverished. He could not bring himself to believe that the human species would tolerate such a system indefinitely, and believed it would soon, and inevitably, find a better way to live.

For a long time after Marx's death it looked as if he might be right. Most intelligent people in the late nineteenth and early twentieth centuries thought so. But those who, in Marx's name, tried to create such a higher form of society only

³ "A Yankee comes to England, where he is prevented by a Justice of the Peace from flogging his slave, and he exclaims indignantly: 'Do you call this a land of liberty, where a man can't larrup his nigger?'" (CW 5:210). In the American south, freeing slaves was seen as what it also actually was, a curtailment of the liberty of slaveholders. Today in the U.S., by the same reasoning as that used by the Yankee in the above quotation, it is universally regarded by the economically dominant political party (the party of corporations and the top 1%) as an encroachment on natural liberty (the liberty of employers, or the freedom of the "free market") when workers seek to organize and bargain collectively, or when the state seeks to regulate working conditions so workers are protected from injury. But in the political world—Marx would say in all class society—there can be no such thing as promoting "liberty" in general. Protecting one person's liberty necessarily requires curtailing another's. The question is only whom we choose to emancipate and whose liberty we choose to curtail. Our monstrously unequal society, like a society based on slavery, systematically favors the liberty of the victimizers over that of the victims. Some egalitarians may try to take account of the point I am making by means of what is commonly called "luck egalitarianism." For example, see Arneson, 2004. I am sure the "luck egalitarians" mean well, and the policies they advocate in practice are probably policies I would also favor. But, to begin with, it is not luck at all, but rather a structural necessity of modern capitalism, that it leaves the majority of workers helpless and at the mercy of capital (see G. A. Cohen in Roemer, 1986). So "luck egalitarianism" (taken literally) seems committed to leave capitalist oppression intact (since this oppression is not a result of luck). Perhaps, however, the idea is that it is luck that determines who is assigned to which class, and the luck egalitarians want somehow to compensate for that. But can they do so, while leaving class society intact? It looks as if "luck" in luck egalitarianism is being understood in such a way that luck egalitarianism requires society to treat people unequally in whatever way turns out to be necessary to make up for the unequal distribution of "luck" (in whatever arbitrary way the luck egalitarian chooses to define this latter term). In that case, I submit that luck egalitarianism is not a form of egalitarianism at all, but rather an open-ended invitation to treat people unequally in whatever ways seem fairer to us than any kind of equal treatment would seem.

replicated the authoritarian political systems that were traditional in their nations and cultures, without even achieving the capitalist world's degree of prosperity. These were never, of course, controlled experiments in socialism or post-class society, since they always operated under the relentless pressure of the capitalist world, which needed them to fail, and did everything it could either to destroy them, or to turn them into self-discrediting engines of backwardness. It might be a point in favor of Marx's vision that it took nearly three-quarters of a century to bring them down. But they left behind them little in the way of encouragement for future post-capitalist experiments. And humanity currently languishes and suffers under the burden of that hopelessness.

Marx also thought that world-historical opportunities made possible by the growth of human productive forces would remain available indefinitely, until we choose to take advantage of them. But in our age of extreme inequality, and threatened with the consequences of climate disruption, there may be grounds for doubting this. Capitalism, therefore, may no longer appear to us as a temporary form, destined to give way to something better. Now it looks like only a quagmire, a quicksand, in which our species, unable or unwilling to extricate itself, may eventually be doomed to perish miserably, or at least to suffer from want, misery, and a greatly reduced standard of living, as well as from even greater economic and environmental instability, all due to the long-term effects—the unsustainable way of life—created by the very technology Marx thought was capitalism's great liberating gift to humanity. At the same time, capitalism has so degraded our humanity—blinding us to the insight of Louis Blanc (or of the New Testament), reducing us to abject slavery to capitalism's inhuman social forms—that the prospect of our extinction may no longer offer any cause for regret—unless we find it in the comfortless words of Willa Cather: “Even the wicked get worse than they deserve” (Cather, 2006, p. 206).

Old ways, traditions, and heritages are held up to us as something to be treasured and perpetuated simply because they are ancient and because humanity has not yet found a way around them. But more often they deserve, for just that reason, to be challenged, opposed, then uprooted and transcended. When he wrote the following words in 1762, Cesare Beccaria was referring to his proposal to abolish the death penalty. But what he says could apply to old usages and hallowed traditions of many kinds, whether religious, moral, political, or economic:

If anyone should cite against me the example of practically all ages and nations, which assigned the death penalty to certain crimes, I shall reply that the example is annihilated in the presence of truth, against which there is no prescription, and that human history leaves us with the impression of being a vast sea of errors, in which there float a few truths, confused and widely scattered (Beccaria, 1986, p. 52).

Our ground for hope is that with the passage of time, these truths may gradually come to prevail. Since the eighteenth century, for instance, many people have listened to Beccaria's rational arguments against the death penalty, and now this inhuman practice has been abolished in all fully modern and decent societies; it lingers on only in backward nations (however prosperous) where stubborn prejudices and brutish traditions resist the advance of reason.

Despite the new reasons for doubt that now challenge the Enlightenment faith in a better human future, the rational grounds for that faith have not disappeared. Capitalism is a callous, exploitative, social system that rewards greed and dishonesty, and subjects the many to the arbitrary power of the few. That it represents an advance beyond slavery, but remains akin to it, is shown by the fact that capitalist ideology is celebrated chiefly in the last developed nation in which slavery flourished, and in which it took a horrendous bloodbath, followed by well over a century of hard struggle, to overcome its social remnants. The same customs and attitudes that once stubbornly supported slavery are now largely devoted to preserving capitalism in all its most extreme and inhuman forms. The French Revolution of 1789 led to terror, repression, empire, and the eventual restoration of reactionary royalism. Throughout much of the nineteenth century, it was held up, not without reason, as an example of what is most to be feared.⁴ It was a disastrous failure, judged by the goals of establishing human equality and a representative republican form of government. When asked about the significance of the French Revolution, the Communist Chinese leader Zhou enlai is reported to have

⁴ Kant's attitude toward the French Revolution, often misunderstood, is worth considering here. He was opposed in principle, of course, to violent insurrections against even despotic governments, and he admitted (in 1798) that "the revolution of a gifted people which we have seen unfolding in our time... may be filled with misery and atrocities to the point that a right-thinking human being, were he boldly to hope to execute it successfully the second time, would never resolve to make the experiment at such cost." But Kant thought that the initial favorable reaction of enlightened observers (the consensus of learned audiences throughout Europe)—their "wishful *sympathetic participation* (*Teilnehmung*) that borders closely on enthusiasm... can have no other cause than the moral predisposition of the human species" (SF 7:85). Kant was not necessarily endorsing the Revolution itself, but he was noting a phenomenon that has become familiar to us, I believe in the twentieth century—for instance, in the positive reactions of intelligent people throughout the world to such events as the defeat of Nazism and Fascism in Europe, the fall of apartheid in South Africa and of the Soviet empire, the election of President Obama in 2008, and their negative reaction to the Soviet invasions of Hungary and Czechoslovakia, and the U.S. invasions of Vietnam and Iraq. Kant took his own enthusiastic attitude towards the Revolution, which was shared by most progressive European intellectuals (though it later turned to disillusionment by many of them), along with our very tendency to deplore the moral corruption of modern society, to be the most hopeful signs we have—very early, as he saw it, in a long, slow process—that the human species may be progressing in history (SF 7:85–9; Anth 7:333; cf. MA 8:116–18).

replied: “Too soon to say.”⁵ There is no reason why the same might not be said of the miscarried twentieth-century attempts at socialism.

The main reason for hope regarding the gradual reform and eventual abolition of exploitative social relations therefore remains the same for us as it was for Marx. It is the thought that surely our species, endowed with intelligence and reason, cannot remain indefinitely so vile and stupid that this is the best it can do for itself. We human beings still exercise some degree of control over ourselves, our world, and our future, and we do not (we cannot) know that our hopes are doomed to failure, any more than we can know that they are destined inevitably to be crowned with success. Capitalism and its associated evils—the crying inequalities attaching essentially to its social relations, the heartless attitudes it presupposes and perpetuates, the environmental disasters it threatens—these remain the chief obstacles in the way of the realization of our hopes. We still have before us the task of finding a way to a higher future, and we still require rational faith in order to undertake it.

⁵ The remark may be apocryphal, perhaps mistranslated, and it may have referred not to the French Revolution of 1789 but to the student uprisings of 1968 that ended the rule of Charles de Gaulle (McGregor, 2011). The wisdom of the imagined remark, however, in relation to the revolution of 1789, is genuine. That is why it is so often quoted.

Coercion, Manipulation, Exploitation

Human beings are social creatures. This means that people are dependent on other people for nearly everything they need in life. That entails, in turn, that they regularly need to get others to do things they need or want them to do—or to refrain from things they need them not to do or don't want them to do. (When, in the following, I speak of needing, wanting, and getting other people to do things, I understand also cases where we get them *not* to do something.)

People get others to do what they want in a variety of ways. Some of these are morally unproblematic, at least most of the time. We can get other people to do things, for instance, because these others have a legal obligation to do them. Or they may have, and respond to, a (non-coercible) moral obligation, or just do it for us out of moral decency. Or we can offer them some inducement sufficient to get them to do what we want: buying and selling, when the buyers and sellers are free and the market is considered fair, falls under this heading. Even if they owe us nothing, and we can offer them nothing, we can also ask favors of people, and sometimes they freely agree to what we ask.

Other ways that people get others to do things according to their needs or wants are morally problematic. It is one of my theses—not one for which I will argue directly, but one that I believe will emerge in the course of this chapter—that when getting others to do what you want is morally problematic, this is not so much because you are making them worse off (less happy, less satisfied), but instead it is nearly always because you are messing with their *freedom*—whether by taking it away, limiting it, usurping it, or subverting it. By “freedom,” I mean the capacity of a rational human adult to govern his or her life, rather than having it subject to the will of someone else. But now I am getting ahead of myself. So let's back up and start again.

We have a range of concepts to deal with cases where getting people to do things you want them to do is morally problematic. Two of these concepts are “coercion” and “manipulation.” People can be wrongfully coerced to do something for others that they don’t want to do, even something they shouldn’t have to do (such as turning a purse or wallet over to an armed robber on a dark street). They can be deceived or misled into thinking they owe us something when they don’t. We can put them in a position where they find it awkward or embarrassing not to agree to accept a proposal or to do us a favor. Or we may play upon their fears, desires, and weaknesses to get them to do things we want them to do. These are not cases where they are forced or coerced to act as we want them to act, and we may not be violating any right they have against us (still less any law), but the way we have treated them may still seem morally wrong. My aim in the following discussion is to clarify the concepts of *coercion* and *manipulation*, and also to relate them to a third concept belonging to the same family, the concept of *exploitation*.

12.1 Moralized and Non-moralized Concepts

I have just spoken of “the” concepts of coercion, manipulation, and exploitation, but I should begin by admitting that people may have different concepts corresponding to these words, for which different philosophical accounts might be appropriate, depending on the problems or issues we want to address. My aim here is to explicate these concepts in a way that helps us think critically about cases where getting other people to do things is morally problematic.

One question that is bound to come up in connection with all three concepts is whether to employ what I’ll call a “moralized” version of them. By this I mean an understanding of the concept in which it is taken for granted as part of the very meaning of the term that what it refers to is immoral. This is the way “murder” is understood, for instance, when the term is treated as equivalent to “*wrongful* homicide.” Some acts of homicide may be justifiable, but to call an act “murder” (when “murder” is a moralized concept) is to say already that it is wrong. It would be self-contradictory to call a homicide “murder” unless you are taking it for granted that it is wrong, so that to show that an act is murder, you must first show that it is wrong. Likewise, a moralized version of “coercion,” “manipulation,” and “exploitation” would be one that took it to be true, just as part of the meaning of these words, that an act of coercion, manipulation, or exploitation is wrong or immoral.

It is certainly possible to understand these concepts in a moralized way. And some philosophers do. Alan Wertheimer, for example, uses a moralized concept of “exploitation” in his book *Exploitation* (Wertheimer, 1999). This is alright if he

wants to examine cases that we assume going in are immoral, and to explicate, under the term “exploitation,” what it is that we consider immoral about them. This usage of “exploitation” might even enable us to treat problematic or controversial cases—for example, capitalist wage labor—by trying to decide whether they are really cases of “exploitation” at all (in the sense of the word in which exploitation is by definition wrongful or immoral). Kant tends to use moralized concepts—for example, of murder (as *homocidium dolosum*) or lying (as *falsiloquium dolosum*)—in thinking about duties and their application. Not all cases of homicide or false speaking need be wrongful, but duties prohibiting certain kinds of actions can be thought of as falling under moralized concepts, which it is the agent’s task to apply with good practical judgment to a particular set of circumstances. Thus all cases of murder or lying can be regarded as wrong, but whether *this* case of homicide is a murder, or *this* case of false speaking a lie, might be the point at issue. This is a perfectly defensible way of thinking about general duties and cases that might be exceptions to them. (And it would permit a philosopher to subscribe to unexceptionable moral rules without being guilty of the inflexibility with which Kant is often charged—sometimes *justly* charged, but only here and there, and not systematically or in principle.)

But we may want to look at morally questionable cases in a different way. We may want to begin by trying to argue that capitalist wage labor is exploitative, whatever our initial moral stance on it might be, and then consider arguments that we should come to consider capitalist wage labor as wrongful or objectionable *because* it is exploitative. If we think that moral argument should proceed not merely by invoking our pro- or con- sentiments, or appealing to our unargued intuitions, but instead by identifying objective facts about a situation that give us good reasons for condemning or approving certain things, then we would usually do much better to use a non-moralized sense of words like “coercion,” “manipulation,” and “exploitation,” a sense in which these words can be used to refer to such objective facts.

I also think that using all three terms in a moralized sense distorts their common usage in certain ways that may mislead us, since I believe all three are quite commonly applied to behavior we *do not* regard as wrongful or immoral. Thus, in my opinion, people can be *rightfully coerced* (under the law) not to violate the person or property of others; a speaker at a public gathering might be admired for *skillfully manipulating* an obstreperous heckler into sitting down and listening respectfully; and a baseball manager may be congratulated for having *exploited* the base-stealing abilities of his leadoff man. To use the words “coercion,” “manipulation,” and “exploitation” in a non-moralized sense does not, however, commit us to regarding these (or any) actual cases of coercion, manipulation, or exploitation

as morally justified. We might use these words in a non-moralized sense while still thinking that coercion, manipulation, and exploitation are always wrong (even wrong in the cases just cited). Some people think that abortion is always wrong, but they don't think that this is true simply by virtue of the meaning of "abortion." They realize that they are using the word "abortion" in the very same sense that it is used by people who think that abortion is not wrong. This, in fact, is what enables us to use the word "abortion" in articulating substantive disagreements over whether abortion is always wrong, or sometimes wrong, or never wrong. For the same reason, if we want to think critically about problematic cases of coercion, manipulation, or exploitation, we might do better to use the non-moralized concepts.

Specifically, we do better with non-moralized concepts if we are considering cases where we are quite sure these concepts apply, but unsure, even so, whether we think the behavior is morally objectionable, or if we want to argue that an act is morally objectionable *because* it has the property of being coercive, manipulative, or exploitative. If we use these terms in moralized senses, all we can make of the claim that an act is wrong *because* it is coercive, manipulative, or exploitative is to take that claim as serving to categorize the *kind* of wrong under which (according to our attitudes, intuitions, or independent arguments) we have already decided the act falls. But this is not the only, or the most interesting, meaning of such claims of the form "...wrong *because*..." The most interesting such claims involve saying that an act has the moral property of wrongness because it has the objective, non-moral property of being such-and-such, and that its being such-and-such gives us, under these circumstances, a *reason* for condemning it as wrong.

12.2 Coercion

There are a number of different philosophical accounts of what it is to coerce someone. A good survey and discussion of them is provided by Scott Anderson (2011). He points out the ways in which different analyses are designed to treat specific questions—for instance, about the agent's lack of responsibility for coerced actions, the assumed wrongfulness of acts of coercion, the way coercion deprives people of freedom or autonomy (in some relatively thick sense of these terms), or the way that coercion relates to political institutions and the law. Compared to most of these accounts, however, my approach may appear very simple and unsophisticated. It proceeds unashamedly from the standpoint of the person who is coerced, or feels coerced, and the account tries to say when, and why, such a feeling is correct. I should therefore say up front that my account of coercion does not serve very well to settle questions about when coercion releases the agent from

responsibility. Nor does it always help us pin the blame on the person who coerces (since as I see it, you may be coerced even if no one coerced you). What this account does, however, is to identify cases of coercion as cases in which your free choices have been in some way or another restricted, removed, interfered with, usurped, or pre-empted. These are, in my opinion, the most basic and important issues there are in determining when someone has been coerced, and in what cases the person is being wronged by coercion. Such issues as responsibility for coerced actions, and who (if anybody) is the agent of coercion, can be considered in this framework, but they turn on matters it does not emphasize.

According to the account I develop here, I am coerced to do something when I either do not choose to do it, or if, when I do choose to do it, this is because I have *no acceptable alternative*. I will speak interchangeably of being *coerced*, *forced*, or *compelled*, taking these terms, whatever the subtle differences in their use, to relate to the same basic concept. (I consider *constraint* as close to coercion, but think your actions or choices can be *constrained* even when you still have a plurality of acceptable options.) Consider a few cases where someone is coerced (forced, or compelled). If I am locked in a room, I am coerced (forced, compelled) to remain there, because I cannot choose to leave. If I give you my wallet because you point a gun at me and threaten to shoot me if I don't give it to you, then I am coerced (forced, compelled) to give it to you because letting myself be shot is not an acceptable alternative. A mountain climber may be forced (compelled) to attempt an arduous climb up the face of a rock, because, due to a sudden change in the weather or an unanticipated landslide, he correctly regards it as unacceptably risky to take any of the other routes available to him. If you offer me a dangerous, degrading job at very low wages, and my only other option is to see my family starve, then I am forced (compelled, coerced) to take it, because letting my family starve is not an acceptable option. I might be coerced (or forced) not to drive your car away either because it is locked and I can't get in, or because I fear I will be arrested and prosecuted for the theft, or (even if I am confident I could get away with it) because I regard stealing as wrong, hence unacceptable. If a judge rules in favor of the defendant because he finds that the facts of the case leave him no alternative under the law, then he is forced (or coerced) by the law so to rule. He might say "I have no choice but to rule in favor of the defendant," meaning that to rule any other way would put his ruling at odds with the law, which is unacceptable.

It might be claimed that if all our actions are causally determined, then we have no choices at all, and if we lack free will, then every action of ours is coerced, constrained, or compelled. I am sympathetic to this claim, but prefer not to open that can of worms in this chapter. Some people think that even if our actions are causally determined, we can still distinguish cases of coercion or being forced from cases in

which we can be called “free” (call this “compatibilistically free”). For the purposes of this chapter, I don’t want to take any position on the issues raised by such claims. This goes along with the fact that my account of coercion is not designed to help us decide when people who are forced or coerced to do something are to be held responsible for what they do, and when they are not. Even those who think our actions are causally determined, and none of them is free, are usually willing to say that sometimes people are wronged because their *freedom* (in some sense of that word) is wrongfully restricted. My discussion of coercion should be relevant to anyone who is willing to entertain this possibility, whatever they may think about issues of free will and responsibility. If someone thinks that the very concepts of right and wrong are illusory, then they won’t be interested in these issues anyway.

12.3 What is an *Acceptable* Alternative?

As these cases are meant to illustrate, the notion of coercion I am explicating here is no clearer than the notion of what is an *acceptable alternative*. I can imagine someone thinking that this makes it so unclear as to make the concept of coercion useless, or at least problematic in making moral judgments. But we should have known already that the concept is problematic, and we should also accept the fact that even so, the judgment that someone has been coerced is often morally important, or even crucial in determining whether some action is wrong or whether someone’s rights have been violated. Someone who thinks such determinations can always avoid using problematic concepts, or that it is the task of philosophy to render all moral concepts totally unproblematic before using them to make important judgments, had better stop aspiring to the unattainable and start living in the real world.

Alternatives can be acceptable or unacceptable for a variety of reasons. Some alternatives might be unacceptable because they threaten an evil so extreme that I can’t or won’t consider it (being shot, letting my family starve), while others might be unacceptable for moral or legal reasons. The standard in all these cases, as I understand it, is not merely subjective. Someone’s just feeling that an alternative is unacceptable is not enough to make it unacceptable. Rather, I am supposing that there is sometimes an objective fact of the matter that certain alternatives are (or are not) acceptable to a given agent under specific circumstances. The application of the term “coercion” usually depends on that fact. Agents can be wrong in thinking that an alternative is acceptable or unacceptable.

Kant distinguishes duties of “right” (*Recht*) from those of “ethics” (*Ethik*) by saying that the former admit of external coercion (*Zwang*), whereas the latter duties admit only of “inner” (or self-) coercion (or constraint), where agents force

themselves to do what they ought through the thought of duty (MS 6:219–20, 379). Ethical duties do sometimes exercise coercion on us, in the sense I mean here, even if there are no external authorities with the right or ability to punish us for violating them. In the case of Kantian narrow or perfect duties, these would be cases of *self-coercion*. In cases of wide or meritorious duties, there is also *self-constraint*—a choice, even where there is a plurality of acceptable options, to act on moral grounds (rational desires and feelings, respect, conscience, love of human beings) even in the face of conflicting inclinations, or at least in the absence of any inclination that might motivate doing the meritorious thing. They are cases of “acting from duty” in the sense Kant uses that term at G 4:397–9, and which I have tried to explain in Chapter 1. (Kantian moral virtue is the strength of character to give priority of such moral grounds over others in your actions generally, so that even when self-constraint is not needed, they do not play a merely “back-up” role in your choice of actions.) Even in the case of external coercion, where all alternatives but one have been rendered unacceptable by the threat of what will happen if the agent takes them, the coercion operates through the agent’s choice of the only acceptable alternative over these others. Kant’s distinction may correspond roughly (and ignoring some of the subtler features of Kant’s theory of merit) to the distinction we make in ordinary life between “the (morally) right thing to do” and “a (morally) good thing to do.” When we say some action is “the *right* thing to do,” we imply that every other alternative is morally *unacceptable*, whereas if it is only “a *good* thing to do,” there might be other actions equally good, some of them perhaps even better. In the first case, we need to *coerce* (or compel) ourselves to do the thing, at least if that is the only way it will get done. But in either case, we might *constrain* ourselves to do the action, out of moral considerations (or as Kant would say, “out of respect for the moral law”).

Disputes over whether someone is coerced may turn on what we decide is or is not acceptable for the person under the circumstances. Someone robbing me in my home may say: “If you don’t tell me the combination to your safe, then you will force me to kill your child. I will have no other alternative.” Here the robber is claiming that my intransigence in the face of his outrageous demand is really a coercive act of mine perpetrated against him. This claim is, of course, outrageously false. A decent person would think that of course he has at least one altogether acceptable alternative to killing my child, namely that of leaving my house and abandoning his criminal project of robbing me. As this example shows, moral judgments may sometimes enter significantly into determinations of what is and is not acceptable for the purpose of deciding whether someone is coerced. But I see no reason to think they always do. To regard it as unacceptable to let yourself be killed, for instance, usually does not involve a moral judgment.

Judgments about what makes you worse off relative to some “baseline” (a criterion used in some accounts of coercion) do not seem to me to do a very good job of settling the real issues about when someone is or isn’t coerced (for the best known such account, see Nozick, 1969). Not only is the baseline always up for grabs (for instance, whether it involves moral judgments or only normal expectations); but also, there is always the question: “*How much* worse off than the baseline are you being made?” If the agent is made worse off relative to the baseline, then that might result in the agent’s having no acceptable alternative, but it also might not. Not every harm or disadvantage imposed on an agent—even imposed wrongly or unfairly—results in the agent’s being *coerced* to do what would be needed to avoid it.

The basic problem with “worse off relative to a baseline” accounts of coercion is that “worse off” is a concept relating to *welfare*, while coercion is about *freedom*. Maybe being worse off sometimes amounts to a deprivation of freedom, but not every case of being worse off is a case of being less free, and not every case of being less free is a case of being coerced.

I understand “coerce” as a success verb: I don’t think a person has been coerced, forced, or compelled to do something unless they actually do it, and moreover had no choice but to do it—either because they had no alternatives, or only unacceptable alternatives. If you threaten me, for instance, with a frivolous lawsuit unless I meet some unjust demand of yours, it might be quite an acceptable alternative for me to contest your lawsuit (which may be irksome to do, but not all that difficult), and in that case you have not in fact coerced me to meet your demand. At most, you can be said to have exerted unfair *pressure* on me—even *coercive* pressure, in the sense that it can certainly be said that you *attempted* to coerce me. But you cannot have succeeded unless I succumb to your threat—and judge the other alternative to be unacceptable. Even then, someone else might correctly judge that I had not really been coerced, if they correctly judge that the option of contesting your frivolous suit was also an acceptable one.

Having your alternatives *narrowed* does *sometimes* raise issues about freedom—when this involves having options removed or rendered unacceptable. But the issue there too is always the presence or absence of acceptable options, and not being well or badly off relative to a baseline. And if you have a plurality of options left, it still does not constitute coercion. Having your alternatives *adversely affected* does raise issues about your *rights* in those cases where worsening the alternatives deprives you of something to which you have a right. But it does not by itself limit your freedom, much less constitute coercion. We will see later that rendering an option less attractive (without either removing it or rendering it unacceptable) is one of the ways of influencing another that falls under the heading of *manipulation*, which is to be contrasted with coercion.

Agents sometimes—unwisely, even disastrously—choose an unacceptable alternative over an acceptable one. Some agents are also faced with choices (commonly referred to as “tragic” choices), in which no alternative open to them is acceptable. The language of coercion may be used in describing these cases: we can speak of someone’s having been “subjected to coercive pressure,” even if they do not do what this coercion would have forced them to do; and we say that the person in a tragic situation was “forced to make an impossible choice.” But talk of coercion is most appropriate when the agent has exactly one acceptable alternative, and takes it. It is especially appropriate in cases where there are other imaginable alternatives that have been either removed or rendered unacceptable for the agent, whether through circumstances or human agency. This last feature of most coercive situations may be what people really have in mind when they describe coercion in terms of someone having been “made worse off” relative to an imagined “baseline.” But for reasons already given, I do not think that is the right way to think about it: I don’t think it’s really coercion unless all the alternatives but one have been rendered *unacceptable*.

There may be different *species* of coercion (moral coercion, physical coercion, through bodily restraint, coercion through threats, etc.) But I take them all to be species of coercion in the sense that they limit, remove, or interfere with an agent’s *freedom*. The freedom to make choices for yourself, rather than having them made for you, is what I take coercion to be most fundamentally about.

12.4 Coercion, Wrongfulness and Responsibility

Clearly, *acts* of coercion (acts that put a person in a position where they are forced, compelled, constrained, coerced) are often wrong. It is wrong for the robber to coerce me to tell him the combination of my safe by threatening to kill my child. But coercion, as I understand it, can also be justified, or even morally required. If a policeman becomes aware of the robbery taking place in my home, then he is forced or coerced (because it is his strict duty as an officer of the law) to apprehend the robber and then, further, to coerce him to desist from his criminal act; the policeman might do this by pointing a gun at him and leaving him no choice—no acceptable alternative—except to get in the squad car and be taken to jail. We normally think the coercion under which the policeman acts, and the coercion under which the robber acts, are here entirely right and just. Under the circumstances, wrong would occur (the policeman would be acting wrongly, the robber would be permitted to act wrongly) if they were *not* coerced (compelled, forced) to act as they do.

Rousseauians may want to add that the judge and the policeman, in obeying the law and doing their duty, are really only following their own will, and that this kind

of rational coercion or self-constraint is not ultimately a curtailment of their freedom. That may be correct, but I don't think the deeper Rousseauian point can be made without including the idea that they genuinely are, in a straightforward and literal sense of the terms, coerced or constrained. That's because the deeper point is that although coercion is always a curtailment of freedom, it may also be true that some kinds of freedom may not be possible without coercion. What Rousseau means, after all, is that there are important kinds of freedom available to you *only by way of your being forced*—to be free. Some freedoms are possible only through the limitation of others. This is a profound truth, which we reject or neglect at our peril.

Does coercion require a coercer? Marcia Baron has suggested to me in discussion that "coercion" implies a coercer, and that in many of the cases I regard as falling under coercion, it would be more natural to say that I am forced than that I am coerced. She may be right about these matters of ordinary usage. But in that case, the more interesting concept is the broader one—being forced—and the narrower one, being coerced by someone, should not get as much attention as it does. My agency is limited, pre-empted or taken over just as much in cases where I am forced, as she would put it, as it is in cases where I am coerced by someone. For that reason, I am also going to speak (I admit perhaps a bit unidiomatically) of a person being "coerced" by circumstances, meaning that they are forced (compelled) by circumstances.

Many cases where I am forced to make a choice, because no other acceptable choice is available to me, are also cases where someone else, as a result, is in a position to coerce me (in the sense Baron would prefer to use "coerce"). This is conspicuously true, for instance, in the case where I am forced to take the job you offer me because my only alternative is to see my family starve. This puts you in a position to coerce me in various ways, simply by threatening to fire me. Baron thinks we should also use the term "coerce" in a moralized way, either as implying wrongness or as implying a moralized "baseline" relative to which a person is or is not said to be coerced. I reject this for the reasons already stated. I repeat that I do not think there is only one possible approach to notions like coercion. For certain purposes, a moralized concept, involving a coercing agent may be superior to the one I am using. I have already stated why I think my approach is the right one to use if we are to raise the basic issues about coercion as the limitation on or pre-emption of free agency.

For these reasons, moreover, I consider it an important thesis, connected to the concept of coercion that I am developing here, that *you can be coerced to do something without there being anyone who coerces you*. The policeman might fear the displeasure of his superiors if he does not arrest the robber, and the judge might fear being reversed by the appeals court. But neither fear, nor the possible

agent involved in it, would have to play a role in their correct determination that they have no choice but to do what the law requires. If I am forced to accept the low-paying, dangerous job you offer me, neither you nor anyone else need have made it the case that if I do not accept it, then my family will starve. Of course, the terms of my employment will permit *you* (subsequently) to coerce me in various ways, as by threatening to fire me if I do not comply with your demands. Or my global subjection to your coercion may be part of the labor contract, as in cases of bonded labor. But my choice to accept your initial offer may be a coerced choice simply because I have no acceptable alternative, even if no one coerces me to make that choice.

That there can be cases of coercion without a coercer seems to me important, if we think there are certain ways in which people have a *right* to be free of coercion. Perhaps the state has a *duty* to prevent anyone from being put in a position where they have no acceptable alternative to taking certain kinds of degrading, dangerous, or underpaid employment. In that case, it should not matter whether anyone put them in this position. Also, having your family face the prospect of starvation might be seen as an issue only about *their welfare*; but looking at things the way I am suggesting enables us to see that it is just as much an issue about *your freedom*. I think this way of looking at things is the right one. Assumptions about coercion and freedom that blind us to it to are responsible for some common and pernicious errors—for instance, that you aren't coerced unless some assignable person coerces you.

Kant holds some doctrines, too seldom appreciated, that help at this point. He thinks that you can violate right (*Recht*), “the right of humanity,” without violating the rights of any assignable individual. This is what is really going on in the notorious example of the “murderer at the door.” Kant is there considering a category of statement (which he calls a “declaration”), on whose truthfulness humanity in general is entitled to rely. If you make a lying declaration to the intended murderer, you do not violate his right (since by his wrongful intent he has forfeited it), but you might still be acting wrongfully. (I have discussed this issue at length in Wood, 2008, Chapter 14, section 2.) Conversely, however, Kant also thinks that your rights can be violated even if no assignable individual violates them. This is the category of “general injustice”; the state may, for instance, have an obligation to tax the rich to support the poor when their poverty itself constitutes a wrong, even if no one has done anything wrongful to deprive them of what is rightfully theirs (Wood, 2008, Chapter 11, sections 1–2). To acquaint yourself with the kinds of cases that make me want to speak of coercion without an assignable human coercer, see also Satz, 2010, Chapters 4–10. Circumstances can force desperately poor women into prostitution under conditions where they are likely to be forcibly

raped, injured, or even murdered by pimps or johns; circumstances force desperately poor parents to subject their children to child labor under deplorable conditions; they also force people into labor contracts that submit them, and sometimes even their descendants, into lifetime bondage. I want to consider these to be cases of coercion, even if we can't locate anybody whose intentional agency resulted in people being so desperately poor that they were forced to do these things. A vital truth, insufficiently appreciated, especially in American society, is that extremes of wealth and poverty not only raise issues of welfare or flourishing, but even more basically, they raise issues of freedom and servitude.

The concept of coercion I am developing may appear to confuse freedom with power, and being coerced with lack of power. There are always going to be endlessly many actions not open to me as long as I am not omnipotent. Notoriously, even an omnipotent being cannot choose both to remain omnipotent and to create a rock it cannot lift. It might seem that the account of coercion I am developing would require us to say such things as that I am *coerced* not to jump to the moon, or square the circle, or be immortal. For here too there seems to be no coercer, but many things I cannot choose to do. Do we really want to say that I am forced (or coerced) not to do all the endlessly many things that lie beyond my power to do? Or that I might be wronged by lacking the power to do them?

Of course not. But I *do* want to say the following things: that I *have no choice* in any of these matters, that my human condition forces mortality on me, that (leaving space ships, airplanes, and hot air balloons out of the story) the law of gravity *compels* me to remain pretty close to the ground, and that the laws of geometry put *constraints* on what figures we can construct, and what geometrical operations we can perform. I would speak of these cases pretty much in the same way I would speak of being forced to stay in the locked room. But let's keep in mind that talk of coercion, constraint, compulsion, or being forced is most natural when there are options open to me that, however, I cannot take because they are *unacceptable*. This is not true of things I merely lack the power to do—jumping to the moon, squaring the circle, or being immune to mortality. For these never were options, and therefore they aren't options I don't (or can't) choose because they are unacceptable. As for the omnipotent being, I think it is true that if this being regards it as unacceptable for it to cease being omnipotent, then it is forced (coerced, compelled) not to create a rock it cannot lift. But unless it does create this rock, it is still omnipotent, and its omnipotence is even constituted in part by the power (which it always has) to create the unliftable rock it chooses not to create.

Some coerced choices (e.g., the policeman's choice to apprehend the robber, the judge's choice to rule for the defendant, even my choice to take the job you offer me) are clearly choices for which the agent is responsible. But some forms of coercion

that remove choice (e.g. being locked in the room, assuming I did not choose to be locked in) as well as some coerced choices (e.g. the choice under duress to divulge to the robber the combination to the safe—let's suppose this time that it was someone else's safe) are not anything for which we would hold the agent responsible. As I have said, my treatment of coercion is not designed to deal with responsibility in cases of coercion. Issues about duress, for example, are very complicated, and far more would need to be said than I can say here in order to deal with the question when people are, and are not, relieved of responsibility for an action because it was an action to which they had no acceptable alternative.

12.5 Manipulation

Being manipulated into doing something is different from being coerced into doing it. The two seem to me to form a kind of continuum, with manipulation occupying the subtler end and coercion occupying the more heavy-handed end. The cases where they might seem to coincide or overlap are really borderline cases, where we are not sure how most suitably to describe the kind of influence under which the agent does the thing. In general, I operate within the following parameters: I don't think you are manipulated into doing something if you do not choose to do it, and you are not coerced into doing it if you had some other acceptable choice. "Manipulation" refers to a way of interfering with or usurping someone's free agency that does not limit or destroy free choice, but rather influences it in certain ways that promote the outcome sought by the manipulator. The question is, which ways of influencing choice constitute manipulation?

This is a difficult question, because manipulation seems to me extremely varied in the forms it takes, and perhaps even less of a natural kind than coercion is. It is one of the strengths of Marcia Baron's APA address "Manipulativeness" that it gives us a taste of the variety of kinds of manipulation there are, the subtly different forms they take, and the equally subtle interplay between the forms (Baron, 2003).

I take Baron's article to distinguish three basic forms of manipulation:

1. *Deception*. This includes outright lying to those manipulated, including making false promises to them, but also misleading them without actually misrepresenting anything, such as by encouraging false assumptions, or fostering self-deception that is advantageous to the manipulator's ends, or getting the manipulated person to "view things differently" or interpret the situation in a light favorable to the manipulator's purposes (Baron, 2003, p. 40, 43–4).

2. *Pressure to acquiesce*. This can involve browbeating, wearing down the other's resistance, and making someone agree to something just to avoid further

discomfort or embarrassment. I think that threats, when the harm they threaten falls short of being coercive, should be categorized as this kind of manipulation. Pressure can also take the form of offering inducements, when (as Baron puts it), they give the manipulated person “the wrong sort of reason” for opting in favor of the manipulator’s proposals (Baron, 2003, pp. 40–3).

3. *Playing upon emotions, emotional needs, or weaknesses of character.* This includes eliciting an emotion with the aim of making use of it. Typical emotions used to manipulate are fear, sympathy, a sense of gratitude toward the manipulator, and feelings of guilt if the manipulated person does not consent to what the manipulator wants. Typical weaknesses of character employed for manipulation are vanity and the need for approval—especially the need for the manipulator’s approval (Baron, 2003, pp. 44–5). Though Baron does not mention it, another common character flaw through which people can be manipulated is *greed*, which tends to make people exaggerate the value or the importance of some prospective benefit, and to make willing to take excessive risks when subject to its allure.

Some manipulation makes use of traits or dispositions common to most people, not even necessarily weaknesses, together with situational factors in which these traits come to look like weaknesses. The well-known Milgram experiments, in which a subject is manipulated into administering to someone (what the subject believes are) severe electric shocks, make use of our apparently normal tendency to defer to people in authority, and our reluctance to incur their disapproval (Milgram, 1974). These traits are manipulated through techniques such as hurrying the subject along (offering no time or opportunity for reflection on what is happening) and also the tendency to go down a “slippery slope” (the shocks are increased gradually, so that it is hard for the subject to see where things are leading until it is too late) (see Sabini and Silver, 2005). Another obvious situational factor here is *surprise*: I do not think we need to wait for more experimental results to know that the results would be very different if the subjects were told ahead of time that what was about to be tested is their willingness to violate their own moral principles or practice cruelty out of deference to authority.

What do all these tactics, maneuvers, appeals, and stratagems have in common that make us consider them various forms of *manipulation*? This is not a question Baron answers directly, but her characterization of manipulativeness as a vice (in the Aristotelian sense) may offer us a clue. She thinks that what this vice gets wrong is fundamentally “*how much to steer others*—and which others, and how, and when, and toward what ends; and more generally, how and when and to whom and for what sorts of ends—to seek to influence others’ conduct... The manipulative person is too ready to think it appropriate—or appropriate for

him—to orchestrate things so as to lead others to act as he wants them to; and in those instances where it is not inappropriate to try to engineer such a change, the manipulative person is too ready to employ means that should not be employed” (Baron, 2003, p. 48). Manipulativeness, as Baron sees it, is a form of arrogance with regard to the kind and degree of control one seeks to exercise over the choices of others—and—I would add, to bring out the contrast with coercion—over the choices they make when they do have other acceptable alternatives. Manipulation may, however, operate by making these other options less attractive, without absolutely removing them or making them unacceptable. This is the main reason why, given the inherently unclear and context-dependent character of questions about what is and is not acceptable, there may be borderline cases where we are not sure whether a person has been coerced or merely manipulated. I think that result is just right; we should draw the distinction between being coerced and merely being manipulated according to whether we regard the alternative as having been made unacceptable or merely made less attractive. The manipulative person “steers” the other as a driver steers an automobile. The automobile is already moving through its own internal combustion engine and momentum, but its direction is influenced by the one who steers it. An automobile, of course, does not make choices, and the manipulator steers the person by seeking to influence these choices—but (as Baron puts it) in the “wrong ways.”

I think we should not misunderstand this last phrase in a way that would turn our concept of manipulation itself into a moralized concept; that is, we should consider the ways manipulation influences behavior as in some way distinctive—a way that perhaps makes us *tend* to regard them as morally dubious—yet still without treating these ways as by definition morally wrong or bad. For I think there can, even so, be cases of manipulation of which, all things considered, we morally approve. Baron does claim to be considering manipulativeness “in a moralized way” (since she treats it as a moral vice: Baron, 2003, pp. 39–40). But she at least entertains that alternative view, which holds that “we should be able to say: “It was manipulative, but was it in any way morally objectionable?” Despite her Aristotelian approach, which would seem to require a moralized concept of manipulativeness, she is also interested in the controversy generated by some (especially Sarah Buss) who have argued, in Baron’s words, that “being manipulative is not as bad as it is made out to be” (Baron, 2003, p. 39; see Buss, 2005). That interest pulls her, and it should pull her, as I have argued, in the direction of a non-moralized conception of manipulation. But later Baron describes the features that tend to make this behavior acceptable as those “that make what would otherwise be manipulation not be manipulation” (Baron, 2003, p. 45).

There are two separate issues here: whether certain kinds of behavior are open to objection, and whether “manipulation” is a moralized notion. A person who approves, or at least condones, behavior Baron would find objectionable because it is manipulative could either say it *is* manipulative, but use the concept in a non-moralized way, and then perhaps go on to argue that it is not objectionable, even though it is manipulative, or else they could accept the moralized use of “manipulative” and deny that the controversial behavior should be called “manipulative.” I’ve already given my reasons for preferring the former alternative, at least for my purposes here.

What is characteristic of manipulative behavior, I suggest, is that it influences people’s choices in ways that circumvent or subvert their rational decision-making processes, and that undermine or disrupt the ways of choosing that they themselves would critically endorse if they considered the matter in a way that is lucid and free of error. Deception by *flat out lying* does this in an obvious way. It feeds the person false information, on the basis of which they make choices they presumably might not have made if they had known the truth. Other, more subtle forms of deception—misleading, encouraging false assumptions, fostering self-deception—do this in more devious ways.

Pressuring gets people to make choices they would not have made if they hadn’t been pressured, choices they might later say they should not have made. (And even if they don’t say this, they might say it was objectionable to have been pressured into it.) Here it is significant that the “inducements” considered under the heading of “pressuring” offer the person “the wrong sort of reason”—a sort of reason that the person would not endorse on reflection, if behaving rationally and operating with normal, healthy motivations intact.

Appeals to emotions, needs, or character flaws also count as manipulative because they too subvert the rational self-government of the person. Such an appeal, as Billie Holiday says of love, can, like a faucet, be turned off and on (by the manipulator), and it also “makes you do things that you know is wrong.” “The virtuous person,” Baron says, who gets right what the manipulative person gets wrong, “tries to reason with the other, not cajole or trick him into acting differently” (Baron, 2003, p. 48).

But we should not misunderstand this talk about reason, taking it in too simplistic or one-sided a way. We rationalists are nearly always careful to avoid this, but anti-rationalists often do it, because it appears to help their (unjust) cause. Not all emotions offer the “wrong sort of reason”: although there is a lot of truth in what Billie Holiday says about love, what is done out of love need not be the result of manipulation and is quite often what the person most fully and rationally chooses to do. (Love is a very heterogeneous and multi-faceted emotion.) The main reason

we should avoid a moralistic concept of manipulation, in fact, is that people can also be manipulated into doing what is objectively rational, and just what they ought to do. At the same time, this can be manipulation that is as morally objectionable as any, as when it is not the manipulator's business to see to it that the other person behaves rationally or does what they ought. Not all appeals to emotion, or even playing on a character defect, need involve manipulation, only those that result in behavior the manipulated agent has good reason to regret, and are encouraged and made use of by a manipulative person for the manipulator's ends.

What manipulation seems always to involve is the circumvention or subversion of a person's capacity for self-government. And this seems to go to the heart of what makes it morally objectionable, when it is. But it also explains why manipulation may be morally acceptable in some cases. As Baron points out, manipulative behavior directed towards children is often used in parenting, and may even be an indispensable part of education, as well as of literary style and other forms of art (Baron, 2003, pp. 45–6). Manipulation is acceptable in dealing with children because children are not yet fully self-governing beings, so there is less rationality there to circumvent or subvert. Manipulation may be acceptable, or even admirable, when it is used to train the child in good habits, or encourage emotional responses that are conducive to the acquisition of the rational and emotional capacities of a healthy adult. (*Reason and emotion are not opposites: emotions—even irrational ones—always have some degree of rational content, and healthy emotions are indispensable vehicles of rationality. We rationalists have always known this; its tediously predictable denial is one of the sad errors of those who reject rationalism.*) We object to manipulating people's emotions mainly when we think the manipulation is not being used in their interest, or when we think it is likely to pose an obstacle to their acquisition or exercise of capacities to govern themselves rationally. Even in the acceptable cases, however, we can see why the techniques of manipulation circumvent or subvert rational self-government. They offer people “reasons of the *wrong* sort,” even when these reasons are used in the interests of leading them forward, or back, to governing themselves according to “reasons of the *right* sort.”

I may not agree with Baron when she suggests that one factor that might justify manipulative behavior is “the worthiness of the end”—at least if her claim is that the worthiness of any end, taken by itself, is sufficient to justify manipulation as a means to it (Baron, 2003, p. 46); but I don't think what I am objecting to is quite what Baron meant. That's because her example of the worthy end is one where manipulation gets someone to see things (grief over a personal loss) in a better—or at least healthier—light. The end is to benefit the person manipulated, and even to benefit him in the direction of helping him to regain his capacities to take control over his life and lead it in a healthy and happy way. A worthy end

might justify manipulation if it steers a child in the direction of self-government, or helps an adult who has left the path of healthy self-government to find their way back. This is why it might be admirable for the public speaker to manipulate the obstreperous heckler into respectful silence. The heckler's behavior may already be less than rational; if it is ill-considered and even discreditable to himself, then the speaker's manipulation may steer him back to a course that a reflective and rationally self-governing person would have followed without needing to be steered.

I also don't think I agree with Baron that manipulation is necessarily made less objectionable if the end is not self-serving on the part of the manipulator (Baron, 2003, p. 47; again, I am pushing the envelope. Baron does not say it is *necessarily* made less objectionable. I suspect her view is that *sometimes* it could be made less objectionable, and I don't disagree with that). People can be manipulated in very objectionable ways by arrogant manipulators for ends that do not benefit the manipulators personally at all. People can even be manipulated by officious moralists to do the what is morally right, and this arrogant use of manipulation might be far more reprehensible than any of the petty moral faults of which the same people would otherwise have been guilty. What matters most, I would say, is the way the manipulation undermines and demeans the person manipulated, by violating and disrespecting their rational capacities to choose for themselves how to live.

12.6 Advertising—and Manipulation Without a Manipulator

Manipulation takes a pervasive form in modern society in the form of *advertising* (see Baron, 2003, p. 44). Advertisements routinely use many of the techniques of manipulation that Baron describes: they lie so blatantly that anyone who believes what they say, or who even entertains it, even against his will, is almost certain to be misled. (No one, on reflection, ever really credits what advertisements say.) But advertisements do not work on us by having what they say actually believed. Advertisements are geared to encourage various forms of self-deception; they frame people's perceptions in ways that are carefully crafted to benefit the interests of the advertiser. Advertisements use constant repetition to wear down our resistance, to reinforce associations at a sub-rational level, to offer us inducements that are illusory, or otherwise "of the wrong sort." Advertisements notoriously appeal to our emotional needs, our vanity, our need for approval, our self-doubt. They infantilize us in order to have their way with us. Even advertisements that manipulate people for what we might think are good ends—to vote for the right candidates, or to give up smoking—do it in bad ways. Advertising never aims to

convince, only to *persuade*. It corrupts the root of rational communication, precludes the possibility of any free human community.

Baron has suggested to me in discussion that advertisement is not as objectionable as manipulation by friends and acquaintances because, first, no one is really deceived by it, since we do not expect truth from advertisers, and second, because we have no relationship with advertisers, there is no betrayal of trust or destruction of a valuable relationship, as may be present when we are manipulated by acquaintances, colleagues, friends, or loved ones.

Her first point would matter only if it meant that advertisement did not work. But, of course, it does work. It is one of the more notable features of advertising that it regularly succeeds in manipulating us, even though we tell ourselves we do not believe it (and perhaps—who knows?—even when we are correct in telling ourselves this). Perhaps advertising manipulates us without relying on deception, but by using other methods (more like pressuring or playing on our weaknesses), or perhaps there is such a thing as being deceived, or acting as if you were (which is just as good from the advertiser's standpoint), even by something you don't consciously or intellectually believe. I find the fact that we are manipulated even without actually being deceived more appalling than reassuring.

As to the second point, it is true that advertising involves no betrayal of trust in a personal relationship, but I find it downright horrifying that our entire social life should be so pervaded by large-scale and entirely impersonal mass manipulation. This seems to me even worse than manipulation by friends or acquaintances. It is terrifying to realize that people routinely permit themselves to be manipulated in this way by complete strangers, without their putting up any organized resistance to it. Baron's reaction itself testifies that we have even come to take it for granted that people will exercise this kind of control over us without feeling there is even anything (or at least not very much) wrong with it. That is probably the aspect of the situation I find most intolerable.

Advertisers have known for decades (if not for centuries) that people are more easily controlled if we discourage them from thinking for themselves—indeed, from thinking at all—and instead encourage them to act on ill-considered impulses arising from what is most contemptible in their nature. Advertising reckons on our vices and our weaknesses, encouraging and enhancing them. As I've said, I think it does this even when it manipulates us to do something we should do—such as vote for the best candidates or quit smoking. With calculated precision, advertising makes worse people of us in order to take advantage of us. Advertising is manipulation in its purest (that is to say, its worst) possible form. A world entirely without advertisements, like a world without wars, would be an immeasurably better world than the existing one.

I have argued that coercion does not require a coercer. People can be coerced by circumstances for which no assignable person (or persons) is (are) responsible. It might be wondered whether such a thing can happen with manipulation as well. I think it can. Advertising is a case in point, because it is not merely a practice engaged in by specific advertisers, but also a social institution that over time shapes people's habits and preferences in the deplorable ways I have just described.

Moreover, the so-called "free" market itself is such an institution. Decisions made in the marketplace often have important consequences for society as a whole and for other people (a phenomenon economists call "externality"). But participants in the market typically lack knowledge of and control over these circumstances, and the context in which they make decisions encourages them not to think about the issues raised by these consequences. Thus the market itself, as a social institution, encourages people to make decisions that focus only on their immediate preferences, often their whims and caprices rather than their long terms interests, much less the good of society, or of future generations. The market itself, I believe, like advertising, plays a systematic role in manipulating people, encouraging them to focus narrowly on their own lives, and even regarding them, only on the present and the immediate future. It encourages people in the idea that they owe nothing to other people except those (such as their family) with whose interests they are immediately engaged. Through the ways in which it enables and intensifies inequalities of wealth and power, the market affects not only the interests but also the freedom of others, while also making it difficult for individuals to see that it is doing this, and creating the appearance of freedom where there is none. I think the ways it does all this amounts to a form of manipulation, for which no assignable individual or individuals bear the responsibility.

It was one of the great achievements of Adam Smith's theory of political economy that it showed how in modern society the market shaped individuals who understand themselves as independent agents, capable of the autonomous management of their own lives and who related to other market actors on terms of formal equality and self-interest, rather than personal dependency. (On this, see Satz, 2010, Chapters 1–2.) He thus helped us to see how what he called "commercial society" educates people for freedom. But the market also has its baleful effects on human freedom, which I fear have been less often appreciated, precisely because the market itself manipulates us so as to conceal them. When we consider this aspect of the way the market shapes us, the term "free market" should always be heard as an oxymoron, not as a pleonasm. Only a society based on markets, and manipulated by advertising, could be one in which many people regard it as a pleasant and even worthwhile activity to "shop"—to expose themselves to commodities (in a store, or on line) that they know they do not need, but with the clear

expectation that they will come to desire them and be willing to part with money in exchange for them. Is it only an idiosyncrasy of mine that I find this sick and revolting? Or is it that the sickness has become so pervasive that a sane attitude now looks like merely some philosopher's idiosyncrasy?

The ambivalent relation of the modern market economy to human freedom was well appreciated by Hegel (PR §§ 189–98, 241–6). He accepts from Smith and other political economists the idea that the market, and especially the discipline of labor involved in participating in a market economy, involves a uniquely valuable moment of liberation and education (*Bildung*), constituting the kind of subjective freedom that he sees as the greatest glory of modern civil society. But Hegel also sees how the transformation of needs and preferences by the market economy, and the kinds of labor pertaining to modern civil society, and the social inequalities created or reinforced by the market, tend to corrupt people's tastes, degrade their human capacities, and lead to the social wrong involved in extremes of wealth and poverty that are a threat to human freedom and even to the ethical values of civil society itself.

12.7 Freedom as Independence of the Will of Others

I hope we are now in a position to see that there is a common theme here, which joins coercion and manipulation, and makes them different ends of a common continuum. Both involve the way in which you might get others to do the things you want, or things you think good for the agent or for others, by taking advantage of the opportunity to usurp another person's rational self-government and substitute your own choices for it. From this point of view, the value in the name of which we should object to both is what we can call freedom as *non-domination*. This can be distinguished from freedom merely as *non-interference*—the absence of some obstacle (which need not be a humanly created obstacle) to some person or persons doing certain actions. Freedom as non-domination is what some recent writers—Philip Pettit and Quentin Skinner—have called “republican” (or “neo-republican”) freedom (see Skinner, 1990; Pettit, 1997). This is a value highlighted in much of the modern tradition in moral and political philosophy—especially in Locke, Rousseau, Kant, and Fichte.

This is the idea that lies behind Locke's conception of human beings as having a natural right to liberty: “The freedom, then, of man and liberty of acting according to his own will, is grounded on his having reason which is able to instruct him in that law he is to govern himself by, and make him know how far he is left to the freedom of his own will” (Locke, *Second Treatise*, Chapter VI, § 63). It is the basis

of Rousseau's insistence on equal liberty as basic to the social contract (Rousseau, 1997, pp. 49–51). Its clearest statement may be in Kant's claim that our sole innate right is the right to freedom "independence from being constrained by another's choice," together with his "universal principle of right": "Any action is right if it can coexist with everyone's freedom in accordance with universal law, or if on its maxim the freedom of choice of each can coexist with everyone's freedom in accordance with universal law" (MS 6:237, 230). Another dramatic statement of it is found in Fichte's conception of the "relation of right" as involving reciprocal recognition between persons, and the summons directed by each of us to the other to leave us an external sphere of free action (Fichte, NR 3:41–7). The basic idea behind this value is that as rational agents we have a reason to want our actions to serve our ends, rather than ends chosen by others without our sharing them. The philosophers just mentioned think we even have a right to freedom as independence from control by others, at least within the constraints imposed by the fact that others have the same right to this freedom as we have.

Being coerced, as I have developed the notion, does not necessarily involve a loss of *this* kind of freedom, because having no acceptable option does not necessarily involve someone else's having the power to usurp your control over your own actions. This notion of coercion is far too general and capacious for that. But in the real world, being coerced often does put others in a position to control you, depriving you of the choices you need to be self-governing. The employer of a worker who must accept any terms of employment or watch his family starve is in a position to take control of the worker's life. More generally, those who are in a position to coerce others are in a position to deprive them of the freedom of self-government.

Manipulation too, more directly (though more subtly), threatens to deprive us of the same freedom. It does not remove our freedom to choose, but does subvert it. And because it operates outside the sphere of coercion, leaving us formally free to make choices for ourselves, it escapes many remedies for it that we might try to implement through coercion. We can try to prevent advertisements, for instance, from lying too openly and too dangerously, and perhaps we can even prohibit those directed (for example) at promoting nicotine addiction among children. In some countries, television advertisements are restricted to certain fairly lengthy periods before or after programming (as is true in the U.S.), so that they do not require the programming to be specifically designed (i.e. warped, disfigured) so as to facilitate regular commercial interruption.

Nevertheless, manipulation often operates through communicative behavior that is indispensable to the exercise of freedom, so that it would be disastrously counterproductive to try to promote free self-government by coercively preventing

manipulation. There are fairly narrow limits on our capacity, even in principle, to prohibit or coercively prevent what is most deeply objectionable about advertising, namely its manipulative character, without at the same time prohibiting or interfering with communicative and self-governing behavior itself. Perhaps there may someday be a form of education that will teach people how to protect themselves from advertising, or from other forms of manipulation to which people are now hopeless victims. But the modern world does not seem to have made any real progress in that direction. It sometimes seems that modern society has made people even more vulnerable to calculated mass manipulation than they were in pre-modern social orders.

12.8 Exploitation

It remains to say something (briefly) about the third concept—*exploitation*—that I want to relate to those of *coercion* and *manipulation*. (For a fuller version of this account, see Wood 1995, 1997, and 2004, Chapter 16.) To exploit something, in the most basic sense, is to *use* it. Exploitation differs from mere use, however, in that what is exploited is used as part of a plan of the exploiter to achieve some end, in which the exploiter exercises some kind of domination or control over what is exploited. In the case of human beings as exploiters and as the exploited, exploitation involves the use by the exploiter, for some end of the exploiter, of something about the exploited—typically, some ability of the exploited or resource in the exploited’s possession—over which the exploiter gains control through some vulnerability with which the exploited is afflicted, of which the exploiter also makes use in gaining control over that about the exploited of which the exploiter makes use for the exploiter’s ends. Thus there are two elements in “human on human” exploitation: there is (a) the vulnerability of the exploited of which the exploiter makes use in gaining control; and there is (b) the capacity or other benefit of which the exploiter makes use through exercising this control over the exploited.

Sometimes we speak not of this *capacity* or *benefit* as the object of exploitation, but of the *person* (him- or herself) as being exploited. I think we do this when we judge that the control of the exploiter over the exploited is sufficiently global, or the object of the exploitation as intimately enough involved with the very personhood of the exploited, that we can regard the entire person of the exploited as being controlled and made use of by the exploiter. I doubt that there is any precise way to determine in which cases this occurs, though there are clear cases of it, and also a tendency on the part of anyone who objects to a form of human-on-human exploitation to depict it as exploitation of the whole person.

I also doubt that we can determine precise conditions for saying that one person is vulnerable to another—as distinct from saying, for instance, that the first person merely wants something from the second that makes the first person do, quite freely and voluntarily, what the second person asks of them. By the nature of the case, the vulnerability of those who are exploited usually makes them more eager to be exploited, and the exploitation in that sense more voluntary for them, than for the exploiter. So voluntariness cannot be treated as a criterion of non-vulnerability.

But there are clear cases of vulnerability. The victim of a successful blackmailer is obviously vulnerable to the blackmailer who exploits him. Someone in dire need of some medication or other temporarily scarce resource is clearly vulnerable to the price-gouger who provides it at a wildly inflated cost. If accepting the dangerous, degrading and low-paying job you offer me is the only way to prevent my family from starving, then I am clearly vulnerable to you, and you are in a position to exploit my labor. I take it to be a pervasive empirical fact about the capitalist system—a fact usually left wholly unrepresented in economists' formal representations of market relations between employer and employee—that the working class is vulnerable to the capitalist class in this way. And this is no accident: those who own make use of the fact of their ownership by exploiting those who do not own: that is the principal meaning of *property* in capitalist society; it is why Marx made put his attack on capitalist exploitation in terms of an attack on private property. There are, however, exceptional instances of this relation—for instance, in cases of athletic superstars employed by sports teams—in which the employer is the vulnerable party and the employee is the exploiter. So the vulnerability of workers is not a formal property of the employer–employee relation as such; it is merely a systematic empirical necessity conditioning such relations considered generally.

Such practices as blackmail and price-gouging are forms of exploitation that nearly everyone finds objectionable (though even they have had their defenders—see Mack, 1982, pp. 273–84, and Zwolinski, 2008, pp. 347–78). But clearly in other cases exploitation is not always bad. Some vulnerabilities (e.g. the weaknesses in a legal case that lacks merit and deserves to be lost) are *good* to take advantage of. In a competitive game, we try to exploit our opponents' weaknesses, and we expect them to try to exploit ours. Good sportsmanship involves an acceptance of these facts; if you didn't expect, or even want, your opponent to exploit your mistakes or failings, you shouldn't have chosen to participate in the game in the first place. But exploitation, when it operates through coercion or manipulation, and when these too are bad, is often bad because it shares in (or rides piggy-back on) the badness of the coercion or manipulation that makes possible the exploitation.

The worst thing about exploitation is that under most circumstances, it is shameful and reprehensible to make use of other people's vulnerabilities for your own ends, and it is degrading to have your abilities and resources, and especially to have you yourself, made use of in this way. This is especially the case where exploitation involves—as it usually does—the exploiter's exercising coercive power over the exploited, depriving the exploited of their *freedom*—their independence of constraint by another's will. Defenders of capitalism usually represent it as a system of freedom. But their conception of freedom is perversely one-sided. It was quite accurately described many years ago by Bertrand Russell: "Advocates of capitalism are very apt to appeal to the sacred principles of liberty, which are embodied in one maxim: *The fortunate must not be restrained in the exercise of tyranny over the unfortunate*" (Russell, 1926).

The truth is that employers, in purchasing the labor of their employees, purchase fundamentally an *authority* over workers, the supposed right to exercise coercive power over them, to deprive them of their independence of being constrained by another's choice, which Kant called "the innate right to freedom belonging to every human being in virtue of his humanity" (MS 6:237). They exercise this authority for the whole of the worker's working life—and often, as far as they can get away with it, outside work as well.

The most important reason that workers have formed labor unions has not been to increase their bargaining power over employers with the aim of gaining economic advantage (higher wages, purchasing power). It is rather the aim of protecting their freedom, limiting the employer's power over their lives. Karl Marx recognized this as the fundamental reason for workers to organize.

Just as little as better clothing, food and treatment, and a larger peculium do away with the exploitation of the slave, so little do they set aside that of the wage-worker. A rise in the price of labour, as a consequence of accumulation of capital, only means, in fact, that the length and weight of the golden chain the wage-worker has already forged for himself, allow of a relaxation of the tension of it. (*Capital* 1: 618).

The *Communist Manifesto* does not proclaim as the end of the working class merely its material betterment, but rather its *emancipation*. Exploitation is degrading to the exploited because it deprives them of freedom, which is fundamental to their human dignity.

Instead of seeing the vulnerabilities of others as opportunities to be exploited, people should regard the vulnerabilities of others as occasions to help them. Sometimes this help would most naturally take the form of helping them to remove the vulnerability. People who are too vulnerable to others, or vulnerable in the wrong ways, lack the dignity we think people should have. When we see many

people extremely vulnerable, through poverty or other forms of disadvantage (that often accompany, cause, or result from poverty), then if we respect humanity and the rights of humanity, we want to see this vulnerability abolished. Those who do not—who regard the vulnerability of others as none of their concern, or who even welcome it as something to take advantage of—are a disgrace to the human species.

Human vulnerability too is not always a bad thing, however. Good and decent human beings often welcome certain kinds of vulnerability, both in themselves and in others. The right response to some kinds of vulnerability in others is not to seek its removal, but rather to make the vulnerable not regret being vulnerable. To love someone, for example, is to be vulnerable to them; if we truly love, we do not want this vulnerability removed, but rather hope the beloved will see it as an opportunity to behave generously toward us and return our love, rather than as an opportunity to exploit us. The dynamic of human relations between those who wrong others, those who are wronged, and the reconciliation of those occupying these roles, is a clear case where human vulnerability can be seen to be a good thing. If I wrong you, this shows how you are vulnerable to me. But I thereby open myself up to your blame and reproach, and that shows how my wrongdoing itself has made me vulnerable to you. But the repentance of the wrongdoer and the forgiveness of the wronged display these vulnerabilities in a way that, in the end, makes us value them rather than regret them. That is why the dynamic of wrongdoing, blame, guilt, repentance, compassion, mercy, and forgiveness is fundamental to the moral and emotional content of all great religions, and represents what is deepest and most appealing about them.

Once we identify a case where exploitation is occurring, it is always a separate and substantive question what should be done about it. Where exploitation is an almost inevitable result of vulnerability, and vulnerability is not the direct result of identifiable human choices, there may be no obvious way to end it. Our first reaction, especially when we begin to appreciate this last point, may be to direct blame or resentment against the exploiter. But this is often not nearly as appropriate as it might seem, and sometimes it is not appropriate at all.

It is significant, I think, that Marx did not react this way to capitalist exploitation. For example, in the Preface to *Capital*, he writes:

To prevent possible misunderstandings, let me say this. I do not by any means depict the capitalist and the landowner in rosy colors. But individuals are dealt with here only insofar as they are the personifications of economic categories, the bearers of particular class-relations and interests. My standpoint, from which the development of the economic formation of society is viewed as a process of natural history, can less than any other make the individual responsible for relations whose creature he remains, socially speaking, however much he may subjectively rise above them" (*Capital* 1: 92).

It would be a complete misreading of this passage to see in it any stance on the general problem of free will or moral responsibility. Marx's point is that, whatever the state of their free will or moral subjectivity, for capitalists and landowners, as long as they remain such (and Marx does not expect them to renounce their property and live like ascetic saints in the wilderness), it may not be an acceptable option, in practical economic terms, to cease their exploitation. They are, in the terms we have been discussing, forced or compelled to be exploiters, and it is for that reason that Marx finds it inappropriate to blame them.

Although exploiters have greater power than those they exploit, it is not always easy for them unilaterally to put a stop to the exploitation, as long as they cannot avoid interacting with those they exploit. It is not always reasonable to expect those who are in a position to take advantage of vulnerability to forego the opportunity. If they simply sacrifice their resources for the purpose of benefiting the vulnerable (whether by removing the vulnerability or treating it with generosity), this might not only seem unrealistically saintly of them, but it might just as easily be experienced by the exploited as condescending: it might be just another way of lording it over the vulnerable, subjecting them to a different kind of humiliation, even as exploiting (taking advantage of) their vulnerability in a different way. At most, it may be reasonable to hope that exploiters will not exploit on such terms as to perpetuate, or even intensify, the vulnerability of their victims. But it may not be obvious to them how to do this; finding such a path may require a moral creativity for which their social role gives them neither the talent nor the incentive.

Contrary to what we might have thought, therefore, it might be a more effective way of combating exploitation if the task of doing so is left to the weaker party: the vulnerable, those who are exploited. This may be why Karl Marx's recipe for emancipating workers from capitalist exploitation took the form of advice and exhortation directed not at the capitalists but at the workers: "Workers of all countries, unite!" (CW 6:539). Organize and strengthen yourselves, so that you will no longer be vulnerable. Taking that advice would not involve the exploiting class condescending to the vulnerable. (Marx was, on these grounds, among the first socialists explicitly to favor the formation of labor unions. Others—Robert Owen, Pierre Proudhon, Ferdinand Lassalle—either opposed them or saw little value in them.) The only burden we might want to impose on the exploiters is that they don't get in the way, that they don't treat the organization and empowerment of the exploited as an illegitimate encroachment on their prerogatives. When Marx directed his exhortations at the capitalists, or their political representatives, this was usually the form they took—don't oppose the development of the working class—and he argued that this was even in the long-term self-interest of the exploiting classes.

Disregarding higher motives, their own interest bids the presently ruling classes to remove all the legally controllable obstacles to the development of the working class. This is why I have made such a prominent place in this volume for, among other things, the history, content and result of the English factory legislation. For one nation should and can learn from another (*Capital* 1:92).

But Marx was not hopeful that the ruling classes would listen, and expected them to fear and resist worker empowerment. We know all too well that these pessimistic expectations were correct.

Marx's more distant hope was that the working class movement would then change society so that the exploitation of one class by another is not built into the social relations that constitute its economic structure. The moral progress of the human species consists mainly in its efforts to remove some kinds of human vulnerability—the kinds that threaten people's dignity—and to cherish others, the kinds of vulnerability that make the vulnerable glad they are vulnerable. Both efforts are aimed at the abolition of exploitation. For these reasons, it is an important measure of the moral level of a society whether it tolerates or even encourages exploitation, or whether instead it declares total war on exploitation, and strives to abolish all but the most marginal and harmless forms of it. By this measure, the society in which we live—modern capitalist society, and especially American society, insofar as it promotes and celebrates capitalist exploitation—is a highly objectionable social order.

12.9 The Connections

Clearly, not all instances of exploitation involve either coercion or manipulation. People can be vulnerable to others, and their vulnerability taken advantage of, without the vulnerable having no acceptable alternative to being exploited. In many cases (if starvation is the only other option, and it is unacceptable), people are forced into being exploited. But in other cases, it may be merely that, given their vulnerability, being exploited is the least unattractive of the acceptable options. The exploitation is none the less real for that. Nor must exploitation come about through manipulation. Exploiters sometimes do deceive, exert pressure that subverts the rational choices of their victims, or play on weaknesses of character. Not all vulnerabilities follow these patterns, and people who simply have a decisive bargaining advantage over others may exploit their vulnerability without needing to engage in any of the tricks of the manipulator. Most manipulation through deception involves no vulnerability—a skillful liar can create a situation where only a person deficient in epistemic virtues would *not* be deceived. Nevertheless, instances of coercion and manipulation overlap in clear and significant ways with

instances of exploitation. People sometimes have no acceptable option but to be exploited. Fear, credulity, greed, or vanity can be vulnerabilities that manipulators can exploit.

A common theme in cases where these three concepts apply—though, once again, nothing is invariably present in such cases—is the deprivation of *freedom as non-domination*: the removal, pre-emption, or subtle undermining by one person of another person's rational control over their own choices and actions. It is also true that coercion, manipulation and exploitation can involve threats to human well-being or happiness. But I believe—and have tried to make a case for it here—that the most serious moral issues raised by coercion, manipulation, and exploitation are issues about *freedom*.

Concluding Remarks

Ancient Confucian ethics recognized four virtues in those who rule: (1) 禮 [*lee*,]: propriety; (2) 義 [*yee*,]: justice; (3) 廉 [*lian*]: integrity; and (4) 恥 [pronounced *tche*, 3rd tone]: which I am told is best translated as “susceptibility to shame.” This fourth virtue is the most important of all. Without it, the others will inevitably be hollow, false, and hypocritical.

In the United States, which arrogates to itself the right to rule the world—coercing, invading, and killing wherever it sees fit—this fourth virtue is one conspicuously lacking in the population generally, and especially in its political leaders. On the rightward side of the political spectrum, this virtue is not merely lacking; it is repudiated with exultant fury, at least insofar as the virtue might apply to them, or to that in the nation’s history their policies would perpetuate. The political left, for its part, is cowardly and corrupt, unwilling to stand up for what it knows to be true, willing to compromise with greed, lies, and bigotry, willing to pretend, for political purposes, that there might be some truth in blatant lies. On which side here is there the greatest reason for shame?

In the nation’s history, the grounds for shame are also more than sufficient. It is not an accident that the nation that most fervently identifies itself with capitalism is not only the most obscenely unequal of developed nations, but also the nation that incarcerates the largest percentage of its population—and is, in that obvious sense, the most unfree society on earth. The United States was founded on the violent dispossession of the native population of an entire continent. Its economy was for a long time built on the abomination of slavery, which was abolished only through a horrendous bloodbath. The racist underpinnings of slavery then remained in full force for over a century, especially in that region that had to be brutally coerced into accepting the abolition of slavery. Racial prejudices, including the ignorant, short-sighted xenophobic persecution of each new wave of foreign immigrants as it begins to make its contribution to the nation,

continue to exert strong influence on our political and economic life even now, and for the future. Even after slavery was abolished, the tradition of willing acceptance by many of the oppression of slavery transitioned too easily into the widespread acceptance of what is most oppressive about capitalism, and that it still incarcerates a disproportionate percentage of its dark-skinned citizens. What is most shameful about the continuing structural racism of American society is not merely the diverse political influence it exerts, but the absurd lengths people go to convince themselves that it is something else (“states’ rights,” “law and order,” “free enterprise,” “excessive government” . . .)—the list of masquerades is subject to variation over time, but the basic political positions, the consequences, the voting patterns, and their ongoing significance, do not change. The lies, pretense, indignant denials, and self-deceptions do less harm than the murder, racism, and oppression they rationalize, but they may be even more disgraceful.

Modern states more generally, at least in the developed world, flatter themselves that they are fundamentally—if imperfectly—free and just societies, and democratic societies, which act in the name of their members, who therefore share in ruling them. Perhaps it is inevitable that people (or at least many people—those who enjoy the privileges, and those afraid to imagine change) should seek to be proud, and not ashamed, of the society they live in and complacent about its faults, even if they recognize that these are serious. Here, the wisdom of the fourth Confucian virtue of rulers seems especially applicable. For the point is not whether there are reasons for being proud of your heritage. There always are such reasons: I am sure my compatriots will hasten to provide them, and to condemn as unfairly one-sided the grounds for American shame that I have offered in the previous paragraph.

It is a corruption, however, to take such reasons for pride as grounds to *exempt* ourselves from shame. We can’t tote up these reasons on two sides of a scoreboard, and regard ourselves as excused from feeling shame if we end up with a positive, rather than a negative net total. I do not say that people may not be proud of what is good in their cultural traditions, only that they combine this with a decent sense of shame for what should not have occurred in the past, and a resolve to change their society so that the same wrongs will not continue. Confucian wisdom did not say that the Emperor, his ministers, his district officials, might not be proud of their rule—insofar as it exhibited propriety, justice, and integrity. Susceptibility to shame is an indispensable virtue in rulers because in those who exercise power and bear responsibility, shame is inseparable from a resolve to change. “Shame,” said Marx, “is already a kind of revolution” (CW 3:133).¹

¹ The shame Marx means, and also that intended by Confucian ethics, is a shame that is *self-imposed* (“Shame,” said Marx, “is a kind of anger directed inward,” CW 3:133). It is imposed through the correct

I am reluctant to urge shame on people from other nations and cultures, but I do feel justified in this regarding my own nation and culture. This is because it is mine, and because it presents itself to me as acting in my name. This gives me the right and the duty to view without illusions the wrongs my society commits, both towards its own people and towards others, and then to criticize, condemn, and seek to correct the wrongs it has committed in the past and that it still commits.

The eighteenth- and nineteenth-century German philosophers discussed in this book take varying attitudes toward the human tendency to take a one-sidedly positive attitude toward one's own culture, and to exempt oneself, and one's society, from shame. Hegel seeks to reconcile us to the social world of the modern state by demonstrating that its kernel (its "actuality") is rational, even if the superficial, transitory husk necessarily contains contingent imperfections. Hegel's argument includes an insightful account of the needs and aspirations of modern human beings; but his attempt to defend the modern state in terms of them is far more doubtful. When we examine the values on which Hegel rests the alleged rationality of modern society—recognition of the abstract rights of persons, the subjective freedom of moral subjects, the harmony of individuality and universality, and of particular self-interest and social good in ethical roles, the participation by all individuals in a greater social good than each one's own private individuality—and then compare these values with the existing institutions of modernity, I believe we are likely to be ashamed of the way the modern capitalist nation state systematically betrays these values, and not much impressed with Hegel's attempt to portray the early nineteenth-century nation state as actualizing them. Thus Hegel's legacy belongs more to the critics of modern society, to those susceptible to shame, than (as he intended) to those who want to defend its essential rationality.

Herder too can be seen as a defender and apologist generally for existing social orders, since he holds, as a matter of principle, that every nation and age should

perception of wrongs you have done or in which you are complicit, where you not only bear significant responsibility for that of which you are ashamed, but even more importantly, where you have the power to do something about it, to improve what lies within the scope of your responsibility, so as to remove or lessen the self-imposed shame. Confucian ethics assumes that rulers have the power to correct and improve the unjust or abusive rule of which they are rightly ashamed; Marx assumes that people have the power to change their society once they see it as inhuman and oppressive. This shame is entirely different from forms of shame that are *imposed from without* on the weak and vulnerable by institutions, authorities, traditional mores, or public opinion—a shame they are powerless to do anything about—for example, the shame inflicted on persecuted minorities, social pariahs, or scapegoats, or on women when they are raped, or even (in some cultures) when they are menstruating. (This includes much of the shame imposed on the poor for criminal behavior, or financial insolvency, or just for being poor.) This latter form of shame is, in all cultures, one of the chief mechanisms of social victimization; susceptibility to it is never a virtue, but rather a terrible misfortune, and also a sign of injustice.

be judged virtuous and happy by the standard of its own way of thinking—which in his view is the only appropriate standard to apply to any age or culture. But Herder makes an exception of his own age's way of thinking, especially insofar as it sets itself up as the judge and the *telos* of others. (There is definitely something in Herder's resistance to that cultural arrogance with which an American can identify.) Herder appears to see the state of European society at the end of the eighteenth century as one of old age and decay; he looks forward to the emergence of a new and higher society in the future. He views the long arc of human history as tending toward *Humanität*—reason and equity for all human beings; he sees different ages and societies as making different (though always imperfect) contributions toward this ultimate goal.

Like Herder, Kant regards humanity as at an early stage in its long, slow, and halting progress toward the fulfillment of its human vocation. He sees the politics, the religion and, far more than is commonly appreciated, the moral attitudes of both the past and present as far distant from what they might be, and should be, if humanity ever lives up to its moral vocation. Kant follows Rousseau in holding that modern society, and its "civilization" has (at least up until now) led more to misery and corruption than to that fulfillment. Alongside "civilization," therefore, Kant points us to a historical process of "moralization," a struggle against the moral corruption of human nature and to the injustice and conflict of modern social life. But Kant is also wary of rapid and violent attempts at social change, fearing that they will lead to despotism, or even barbarism; and he defends (on various grounds, some of them shocking and disgraceful to our eyes now) many of the moral judgments and the social institutions existing in his time that now seem to most of us incompatible with the fundamental principles of right and ethics that he articulated.

Fichte sometimes follows Kant in this last tendency (particularly when it comes to women and the family, though there his views are a rich mixture of the utterly disgraceful and the admirably far-sighted). Fichte also acknowledges, as Kant does, that humanity has far to go, and seems open to a future that he himself (or anyone now living) may be unable to comprehend. He was in his own day condemned and persecuted as a "Jacobin" and a dangerous political radical. And though he finds a way of making his peace with the present political order, he does so only on the following highly doubtful terms: "Any constitution of the state is in accord with right which does not make it impossible to progress toward something better . . . Only that constitution is completely contrary to right which has the end of preserving everything as it presently is" (SL 4:361). This is not too far from Marx's "ruthless criticism of everything existing" (CW 3:142).

As I said in the Introduction, as a scholar and philosopher it has been my first obligation to present these historical philosophers in their own terms. That is why, with only a few exceptions, I have drawn my own conclusions from their thought largely in the Introduction and in these Concluding Remarks. Now I state some conclusions relevant to our own time that I believe follow from a sound consideration of the philosophy discussed in the twelve chapters of this book. I repeat what I said in the Introduction: these conclusions were not all accepted by the philosophers discussed in this book. I also reiterate my explanation for that: these philosophers were too accepting of the customs and practices of their time to draw consistent conclusions from their own doctrines; with longer historical perspective, we are in a better position to appreciate the viewpoint to which their philosophy leads.

I do not offer the following as practical political proposals, but rather in the same spirit as Fichte presented to his audience in Jena a vision of a possible human future:

I am not claiming that things are as I have described them. But according to our practical concepts . . . this is how things ought to be, and we can and should work to make them so . . . We can at least catch a glimpse beyond ourselves of an association in which one cannot work for himself without at the same time working for everyone, nor work for others without working for himself (VBG 6:321).

1 All Should be Left Externally Free to Govern their Own Lives

Individuals should be free from constraint by the choices of others, from external coercion, paternalistic interference, pressure, manipulation—when they do things that affect only themselves, and in their dealings with others, recognized as free and equal, on entirely voluntary terms. This includes what they believe: about morality, religion, philosophy, science; how they choose to inform and entertain themselves—what books they read, what plays, movies, or television they choose to watch. It also includes what substances they put in their bodies; whom they choose to befriend, or to love (including sexually); and how they choose to deal with their own bodies (in matters of life, death, and reproduction).

This same freedom should belong to all adult human beings. The decision to treat an embryo or fetus inside a woman's body as already a person—a being deserving of external coercive protection against the woman of whose body it is a part—is simply to presuppose (as if it were a sort of settled metaphysical fact, and without permitting the question even to come up for debate) that women are

not free human beings—that it is not the women themselves who have final say over their bodies, but instead others—the self-appointed guardians of the alleged fetal person, those who happen to believe in this personhood and who arrogantly demand to insert themselves into another person’s life as the protector of these imagined rights. The self-evident truth, however, is just the opposite: that every adult human being, simply as a free person, has unquestionable rightful authority over the life processes going on in his or her own body. If guardianship is required for the potential or unborn human life inside a woman’s body, she is the sole person who could ever rightfully exercise it—no other private persons, no officious priests or intrusive religious groups, have any business interfering. No person’s rights of conscience or religious belief could ever extend so far as to give them authority over the body of another adult human being. As for the state, it is its chief business to defend, with all its coercive might, the rights of all persons over their own bodies. It should never let itself be complicit in, or the agent of, depriving them of their rights.

Women have for ages been treated as chattels, domestic labor, providers of sexual services, baby-makers. They have been subjected to the authority of others, especially of men, and especially in their sexual and reproductive lives. The only issue relating to abortion is whether these injustices will finally be ended. It is a sick sophistry to present the backward policy of perpetuating them as if this were actually an *extension* of the scope of human rights to a new class of (unborn) persons, like the emancipation of slaves. The analogy between this issue and slavery is wholly apt, but it makes entirely the other way.

Everything falling under self-regarding matters collectively constitutes a sphere in which the external freedom of all adult human beings should be protected, both from the private interference of others and also from the coercive power of the political state. The chief office of the state, in fact, is to protect this very freedom, and to prevent both other private persons and the state’s own representatives from interfering with it. Regarding these matters, however, there may be serious *ethical* issues—non-coercible ethical duties demanding the self-constraint of rational agents. We may not like Kant’s eighteenth-century term “duty”—which sounds stuffy to us and has external institutional connotations that were not part of Kant’s theory—but we recognize claims on us made either by other people, or by moral principles.

We have a moral duty, in this sense, to believe rationally, according to evidence and proof, not forming our beliefs through self-deception, wishful thinking, social conformity, or blind adherence to some political or religious dogma. There may be *ethical* constraints on the taking of drugs, whether we eat meat, which voluntary sexual practices we engage in, when, and with whom. A woman arguably

has a non-coercible ethical duty to value the unborn human life inside her body, and to consider this value in making the decisions that by right are solely hers to make. More generally, there may be moral issues of life and death regarding the life-processes going on inside our own bodies, and duties we owe ourselves regarding them.

On these moral issues we may seek the views of others, or seek to inform them by reasoning with them; they may even have a non-coercible moral duty to listen, and we may even owe it to them, as a moral duty, to reason with them about their ethical duties, as long as they are willing to listen to us. But we must not coerce them, either to fulfill their ethical duties, or to listen to our arguments and our preachings. (If someone is trying to deprive a woman of her right over her own body, for instance, then she need not listen to anything they might say about how she ought to choose to exercise that right. If they want to convince her to choose one way rather than another, they must first establish their credibility by defending her right to make the choice.)

It is also morally questionable what means we may use in our attempts to persuade them: we must always address them with respect as free rational agents, who have a right to reject any pressure or manipulation that would subvert their rational self-government. Above all, the state should never exercise coercion over people in any of these purely ethical and self-regarding matters, or symbolically pay homage to the religion or the moral beliefs of the majority, or of those in power. For the state's task is to speak in the name of all; it behaves wrongfully when it displays allegiance to beliefs or values that are independent of the coercive protection of right.

2 The External Freedom of All Must be Restricted Merely so that All May be Free

In matters that are *not* self-regarding—whenever people's behavior threatens the rightful freedom of others—freedom must be restricted through lawful coercion, and this restriction is justified solely in the name of freedom itself. The issue should never be directly the advancement of the *interests* of well-being of others, but always only of their *freedom*. However, people's freedom is in jeopardy when they lack the resources necessary for a free way of life—enough to eat, clean water, clothing, a place to live, adequate medical care, and social services. Want of these goods makes them vulnerable to coercion and exploitation by others. It is the business of the state, as the protector of rightful freedom, to see to it that no one is so in want that they lack what they need in order to live freely.

The issue is never whether we should “trust” those to whom we have given governmental power. Of course we should not. But when people insist that “our” freedom is threatened by a government that is “too large,” what they really mean is that the freedom of the privileged to tyrannize over the rest of us will be curtailed if government is given enough power (and the right power) to protect the freedom of all. That is always the “too large” government that they oppose. The truth is that it is always the chief business of society acting collectively to protect anyone whose freedom is threatened by the coercive power of others. The power that threatens rightful freedom takes three principal forms—the power of *direct violence*, the power to *deceive* others or to keep them in ignorance, and the power of *wealth and property*.

People must be protected from the threat of direct violence. Private citizens should not be permitted the capacity to kill others with ease. They should not possess firearms, especially firearms capable of being concealed, and firearms capable of killing many people in a short time. At the very least, their access to firearms should be sharply restricted, so that it can be guaranteed to pose no threat of wrongful violence to others. This is not a subtle or complex issue. As with health care, it has long since been solved by virtually every developed nation except the United States. Only brutish and barbarous people could regard unrestricted access to firearms as belonging to people’s rightful freedom. On the contrary, the right that all should have is to be free of the coercive power of firearms in the possession of other private persons, as well as from the arbitrary use of violence by the state. In any sane and civilized society, it would be universally acknowledged that it is everyone’s right to be as free as lawful state coercion can possibly make them from the mortal threat to everyone posed by the private possession of firearms. It is therefore difficult to argue with people from other countries who shake their heads at our strange obsession with deadly weapons and say that in America human life is cheap, and that with us the power to kill takes precedence over the right not to be killed.

Freedom of political speech requires far more than the formal prohibition on state or private interference with the right to say what we think. It demands also that the public be protected from the systematic distortion of the truth and manipulation of the political process that is the inevitable consequence of great inequalities in wealth and power combined with the dissemination of facts, opinions, and arguments controlled by mass media. Above all, individuals should be protected from the infringements on their freedom that are inevitable in an economic system in which the opportunity to work, hence opportunity to acquire the means of survival and activity, is afforded them only through the choices of individuals with power over them and interests adverse to their own.

People should also be protected from fraud and deception in private dealings. Advertising, discussed in Chapter 12, is systematically used to lie to people, to put them in a position where they make market decisions in ignorance, or under the influence of irrational impulses. Abstract economic theories (of the “free market”) usually assume equal knowledge on the part of market agents. In the real world, this assumption is false, and it is deliberately made to be false by those who have the power to make it false because they benefit from its falsity. Much of the U.S. financial system—credit, investment, insurance—is based on secrecy and fraud, as is much of the corporate-controlled consumer sector of the economy. Failure of state regulation of the economic system leads to the systematic violation of the rights of investors, borrowers, and consumers. This is not a mere “imperfection” in a system governed fundamentally by “free market” principles. It lies at the very foundation of the actual capitalist system.

From what has already been said above about the rights of individuals in self-regarding matters, it follows that it is wholly correct (according to the language of Virginia’s HB 1468, 1996) to “legalize intimate sexual acts between consenting adults which are done in private and for noncommercial purposes.” There is, however, a well-known economic argument, resting (certainly invalidly, perhaps even facetiously) on a parallel with this. It holds that the state must not regulate “capitalist acts between consenting adults” (Nozick, 1974, p. 163). The parallel rests on a ludicrous misrepresentation. To begin with, wherever the consent to sexual acts is subject to the same coercive economic pressures as generally occur in the relation between capital and wage labor—for example, where women are forced by need to engage in sex precisely “for commercial purposes”—then the state may rightfully interfere with these sex acts. (But it should always do so to protect the freedom of the vulnerable, rather than as their persecutors, as usually occurs in the morally or religiously inspired coercive laws punishing prostitution). The parallel also fails more generally. For in the real world of capitalism (as distinct from the idealized theories of economists and the pernicious fantasies of capitalist apologists), capitalist acts are truly between consenting adults only by way of exception. The capitalist market place is, and always has been, essentially one in which those who own have a decisive bargaining advantage over those who labor. Laborers are systematically vulnerable, and employers are systematically in a position to exploit this vulnerability. Capitalist labor always has been, and always will be, an essentially unfree form of labor. If the formality of labor contracts could ever take place between truly free and equal economic agents, it would take a radical revolution to bring this about. Defenders of capitalism have never shown any disposition in that direction.

The freedom of the exploited can be protected only by drastically curtailing the freedom of the exploiter to exploit. Rightful freedom for all requires that society,

in one way or another (and often through the coercive state) should behave very differently from the way it does in regard to the economic organization of society. It should regulate economic conditions to prevent exploitation, both by protecting the vulnerable through interference with those who are in a position to exploit, by enabling those who are exploited to strengthen their position by their own efforts, and by redistributing power (including wealth) so as to prevent people from being in a position of vulnerability that others may exploit.

None of this is commonly acknowledged because it pleases us to misrepresent our political and economic institutions through idealized, apologetic formalisms: our political system is democratic, ideally giving every citizen one vote and equal access to political power and political office. Our market economy is free, ideally operating through consensual contracts and exchanges between free agents who address one another as equals. Only in practice none of these representations is the way the world really is. And whenever efforts are made to bring the real world closer to them, these efforts are successfully opposed by powerful interests and ideologies arising directly out of the system itself, often rationalizing themselves through precisely these same false idealizations and formalisms.

Those who are comfortable, resigned, afraid of change, afraid of freedom, are persuaded by the formalism of "the free market" that the wealth people have is always something they have earned, and that whatever hardships or impoverishment people incur they must deserve. Capitalists represent themselves as creative and enterprising, the well-springs of all social wealth and the common good. It is then argued that taxation of wealth, at least when it supports services available to others, is a form of theft, or even as involuntary servitude.² They do not see how much the material of their creations consists of the lives of others, whose labor is available to them through their vulnerability and society's neglect, or its deliberate refusal, to provide these laborers with what is needed to make their labor a choice, rather than a form of servitude. This representation also ignores the extent to which wealth in our society is not the result of beneficial creativity, but instead of devices that add nothing to the common good, but rather defraud investors and consumers, despoil the natural environment that belongs to everyone, and even plunder the state, whose self-protective regulations, inadequate as they are, the

² Nozick, 1974, pp. 169–72, represents taxation on earnings as equivalent to forced labor, or slavery. Taxation, however, is monetary payment, which quite distinct from compelled service. In contract law, the distinction is clearly recognized: failure to provide a contracted service can be compensated by monetary damages, but never by specific performance, which would turn contracts for service into a form of involuntary servitude. Nozick himself, however, remains consistent in equating money with human liberty, and thinking that the latter may be bought in exchange for the former. Thus Nozick's conception of a free society (even his "utopia") includes the possibility that people might sell themselves into slavery (Nozick, 1974, p. 331).

exploiters then depict as restricting their rightful freedom. Meanwhile, an ever smaller minority of the population comes to possess an ever larger percentage of the wealth, and an ever larger majority is subjected to power of the super-rich and corporations “... with liberty and justice for all.”

No one can ever *earn* the exercise of arbitrary power over the lives of others; still less can they rightfully inherit it or obtain it through what has accrued on their investments. No one could, by any means whatever, become *entitled* to the wealth and privilege that gives capitalist corporations the dominance they presently exercise over workers, consumers, and even the state. At the other end, no one can ever *deserve* to lose their freedom through economic vulnerability, whether through lack of talent or lack of ambition, still less through lack of family connections, inferior education, or just through bad luck, and especially not through the workings of a social order stacked against them.

It is not only the right but even the responsibility of the common power to guarantee the freedom of all, and thus to prevent the extreme inequality that inevitably threatens freedom. This is not “government interference in our lives”—as if the state’s protecting the rightful freedom of the vulnerable amounted to paternalistically usurping the decisions of those whom it protects from coercion by others. No doubt the state’s protection of the rights of the vulnerable will involve government interference with the lives of those who would wrongly take advantage of the vulnerable. But that is not paternalism. Those who demand that to which they are entitled by right should not be represented as living off “charity,” just as their “entitlements” (often earned by law through a lifetime of labor) should not be represented as illegitimate greed. The greedy are those who are accustomed to using state power in their own interests, and are used to being exempt, to an egregious extent, from regulation, taxation, or any other state action that protects others from their predations.

The systematic failures of our society to live up to its idealized formalistic self-representations, and the false pretense that it does so—these are not mere “imperfections” in a social and political system that is fundamentally just—as if the idealizations were the fundamental reality, and the departures from them were mere unfortunate contingencies. On the contrary, our society, on issue after issue, deliberately and on principle gets things exactly backward: the state, motivated by moral prejudice and religious superstition, seeks to regulate the self-regarding behavior of individuals, while it treats rights-infringing behavior—in matters of fraud, economic exploitation, even the threat of direct lethal violence—as something that, as a matter of fundamental right, should be completely immune to state interference. Any attempt to remedy these “imperfections,” moreover, is denounced as “socialist” or “communist.” But to reject with horror these labels,

to refuse even to admit the ideals behind them to your universe of possibilities—a world that does this can have nothing good in its future.

3 Hope When the Future Remains in Doubt

As Kant put it: “It is not all one under what title I get something. What properly belongs to me must not be accorded me merely as something I beg for . . .” (Ak 19:145). “He who needs nothing but justice can hold [the unjustly privileged] to their debts and does not need to be submissive” (Ak 20:140–1). About eighty years after Kant wrote these words—but now it is one-hundred-and-sixty-five years ago—the *Communist Manifesto* summed up the chief aim of the working class movement:

In place of the old bourgeois society, with its classes and class antagonisms, we shall have an association, in which the free development of each is the condition for the free development of all (CW 6:506).

Marx and Engels displayed confident revolutionary bravado in assuring their readers that the way forward towards a society of this kind was clear, and also inevitable. But today no honest person can feel such confidence in either respect. Those who, at least a generation after Marx died, pretended to lay claim in his name to his hopes, are now generally thought to have failed miserably; they are not fit objects for imitation. But if they failed to find a way beyond capitalism, it does not follow that none can be found. Still less does it follow that capitalism is a satisfactory social system.

Whatever crimes against humanity Stalinism committed, there were (and still are) good reasons for seeking a way beyond capitalism to a freer and more human way for people to produce and to co-operate. This truth is so obvious that it takes a high degree of stubborn, doctrinaire perversity to want to deny it. The spirit of what is now called (misnamed) “conservatism” in America—its ruthless extremism, hatred of moderation or caution, intolerant dogmatism, refusal to compromise, blind insistence on sectarian ideology to the exclusion of beliefs based on evidence—all this resembles nothing so much as the spirit of dogmatic bolshevism a century ago. Such movements cannot ultimately prevail, but they can last for a long time, and do terrible harm as long as they persist.

Marx has been interpreted in just such a spirit, both by self-appointed “Marxists” and by their enemies, perhaps because this spirit has been something they have shared. But it is doubtful that Marx himself shared it. One of the most constant themes in his writings is that humanity, and the working-class movement must be in a constant process of development. He never tires of repeating that the working-class movement has made mistakes, must never shrink from admitting them, and must constantly learn from them if it is to fulfill its historic vocation.

Surely he was right about that, and what he said applies at least as much to our time as it did to his.

There is no consensus about the ultimate goal; there was equally little in the 1840s. Marx, after all, said that “everyone will have to admit to himself that he has no exact idea what the future ought to be” (CW 3:142). But Marx at least felt a sense of inevitability that some way to a more human world would be found. Yet we can still know that some things in our society are very wrong, unacceptable, and demand to be changed. The first steps in correcting them should be clear enough, even if the ultimate goal is not. We should not consider evils acceptable merely because we do not yet know how to correct them, or permit ourselves to abandon hopes merely because we do not yet know how to fulfill them.

The aim of Hellenistic philosophies was *ataraxia* (freedom from care, freedom from trouble or worry). A similar state of total contentment is the aim of many religions. I hope we have learned enough by now to see that *ataraxia* is not compatible with any honest assessment of the human condition, especially the state of our social institutions. Such a state would be only the pitiful, contemptible, and even reprehensible condition of Mill’s “fool satisfied.” Tranquility and freedom from care is not a worthy goal for philosophy, or for religion, but rather a crime against humanity.

To have seen this is part of the legacy of modern culture. We have seen it in Kant, Herder, Fichte, and Marx, just as we can find it in Mill. In his poem *Faust*, Goethe correctly represented his hero’s redemption as achieved through a never-ending, ever hopeful, but never wholly successful (and always discontented) struggle, ending only in death, against the very evil through which he obtained the power, and the will, to struggle. At the end of the poem, Faust is saved in eternity: in heaven the insufficiency of his earthly strivings count as a completed achievement:

*Alles Vergängliche
ist nur ein Gleichnis.
Das Unzulängliche
hier wird Ereignis...*

*Everything transitory
Is only a parable.
What was insufficient
Has here become event...*

Goethe, *Faust* II, 12104–7

Faust’s temporal struggle, however—his absence of tranquility, through which he earns his redemption—is simply life as it should be lived. There is no point in time—during life—where *ataraxia* is a goal we can decently entertain. Faust acknowledges this in his dying words:

*Das ist der Weisheit letzter Schluß:
Nur der verdient sich Freiheit wie das Leben,
Der täglich sie erobern muß.*

*This is wisdom’s ultimate conclusion:
He alone merits freedom and life
Who must every day reconquer them.*

Goethe, *Faust*, II, 11574–6

The same modern wisdom also comes without the gilding of religious poetry: “No doubt,” said Beauvoir, “it is always more comfortable to submit to a blind enslavement than to work for one’s liberation; the dead, for that matter, are better adapted to the earth than the living” (Beauvoir, 1952, p. 263). There will be no shortage of time later on for a comfortable reconciliation with the existing social world, for complete freedom from all the disturbing emotions: anger, shame, care, doubt, worry, fear, and hope. But we all defer this natural, inevitable *ataraxia* as long as we can; anticipating it ahead of time is only complacency, or resignation—a bad way to live.

Marx’s hopes, along with his critique of modern society, are far from having been discredited by the failures of the movements that flattered themselves they were acting in his name. The common dismissal of these hopes—the acceptance of the thought that they have been discredited—is, I think, in no small part responsible for the fact that we have been, for most of my lifetime, on exactly the wrong, the backward path. The more Marx has been dismissed and forgotten, the more the world has come to resemble the unstable, inhuman system about which he wrote, which he hoped humanity would revolutionize.

If there is to be a truly human future, history must resume the course projected for it collectively by the thinkers treated in this book—towards Kant’s and Fichte’s more perfect condition of right and moral progress, toward Herder’s *Humanität*, towards the free society Hegel tried (with spectacular unconvincingness) to portray as the actualized modernity of his own time, and that Marx (more correctly) perceived to be something that could come about only after revolutionary social changes. If we are to find a way forward to a freer human future, we cannot afford to ignore the thoughts of the classical German philosophical tradition that provided the background for the idea of “an association in which the free development of each is the condition of the free development of all,” a society where “one cannot work for himself without at the same time working for everyone, nor work for others without working for himself.” The chapters of this book, though they deal with different thinkers and various topics, should be seen collectively as an attempt to make a case for that proposition.

Bibliography

Abbreviations for writings by Kant, Herder, Fichte, Hegel, and Marx are provided in the List of Abbreviations.

Other References

- Adams, Robert M. (1999). *Finite and Infinite Goods*. Oxford: Oxford University Press.
- Allison, Henry (1990). *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.
- (2011). *Kant's Groundwork for the Metaphysics of Morals: A Commentary*. Oxford: Oxford University Press.
- Anderson, Scott (2011). "Coercion," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL:<<http://plato.stanford.edu/archives/win2011/entries/coercion/>> (accessed June 21, 2013).
- Anscombe, G. E. M. (1963 [1957]). *Intention*, 2nd edition. Oxford: Blackwell.
- Aquinas, St Thomas (1988). *On Law, Morality, and Politics*, ed. William P. Baumgarth and Richard J. Regan, S.J. Indianapolis: Hackett.
- Arendt, Hannah (1982). *Lectures on Kant's Political Philosophy*, ed. Ronald Beiner. Chicago, IL: University of Chicago Press.
- Aristotle (1999). *Nicomachean Ethics*, tr. and ed. Terence Irwin, 2nd edition. Indianapolis: Hackett. [Cited by Becker number.]
- Arneson, Richard (1989). "Equality and Equal Opportunity for Welfare," *Philosophical Studies* 56: 77–93.
- (2004). "Luck Egalitarianism Interpreted and Defended," *Philosophical Topics* 32: 339–49.
- Baiausu, Sorin and Timmons, Mark (eds) (2012). *Kant on Practical Justification: Interpretive Essays*. Oxford: Oxford University Press.
- Baron, Marcia (2002). "Acting from Duty" in Allen W. Wood (tr. and ed.), *Kant: Groundwork for the Metaphysics of Morals*. New Haven: Yale University Press, pp. 92–210.
- (2003). "Manipulativeness," *Proceedings and Addresses of the American Philosophical Association*, Vol. 77, No. 2 (Nov., 2003), pp. 37–54. URL: <http://www.jstor.org/stable/3219740>
- , Pettit, Philip, and Slote, Michael (1997). *Three Methods of Ethics*. Oxford: Blackwell.
- Beauvoir, Simone de (1952 [1949]). *The Second Sex*, tr. H. M. Parshley. New York: Knopf.
- Beccaria, Cesare (1986 [1762]). *On Crimes and Punishments*, tr. David Young. Indianapolis, Hackett.
- Beiser, Frederick C. (1987). *The Fate of Reason*. Cambridge, MA: Harvard University Press.
- (1992). *Enlightenment, Revolution and Romanticism*. Cambridge, MA: Harvard University Press.
- (2002). *German Idealism: The Struggle Against Subjectivism*. Cambridge, MA: Harvard University Press.
- Bentham, Jeremy (1948 [1789]). *The Principles of Morals and Legislation*. New York: Hafner.

- Berlin, Isaiah (1997). *Three Critics of the Enlightenment: Vico, Hamann, Herder*, ed. Henry Hardy. Princeton: Princeton University Press.
- Bernstein, Alyssa (2008). "Kant on Rights and Coercion in International Law: The Implications for Humanitarian Intervention," *Jahrbuch für Recht und Ethik/International Yearbook for Right and Ethics* 16: 57–100.
- Blanc, Louis (2012 [1851]). *Plus de Girondins*. Boston: General Books.
- Broome, John (2005). "Does Rationality Give Us Reasons?," *Philosophical Issues* 15: 321–337.
- Buss, Sarah (2005). "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints," *Ethics* 115 (January): 195–235.
- Byrd, B. Sharon and Hruschka, Joachim (2010). *Kant's Doctrine of Right: A Commentary*. Cambridge: Cambridge University Press.
- Cather, Willa (2006 [1922]). *One of Ours*. West Valley, UT: Waking Lion Press.
- Churchland, Paul M. (1981). "Eliminative Materialism and the Propositional Attitudes," *Journal of Philosophy* 78: 67–90.
- ____ (1988). *Matter and Consciousness*, revised edition. Cambridge, MA: MIT Press.
- Cohen, G. A. (1986). "The Structure of Proletarian Unfreedom," in John Roemer (ed.), *Analytical Marxism*. Cambridge: Cambridge University Press, pp. 237–59.
- ____ (1989). "On the Currency of Egalitarian Justice," *Ethics* 99: 906–44.
- Collins, Randall (1998). *The Sociology of Philosophies: A Global Theory of Intellectual Change*. Cambridge, MA: Harvard University Press.
- Connell, Francis Joseph (1967). "Double Effect, Principle of," *New Catholic Encyclopedia*, Vol. 4. New York: McGraw-Hill.
- Constant, Benjamin (1964 [1795]). *Des réactions politiques, sixième cahier, No. 1, Écrits et Discours politiques*, ed. O. Pozzo. Borgo: Pauvert.
- Damasio, Antonio (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: Harper-Collins Avon.
- Darwall, Stephen (2006). *The second person standpoint*. Cambridge, MA: Harvard University Press.
- Descartes, René (1964–1976). *Oeuvres*, ed. Charles Adam and Paul Tannery. Paris: Vrin. [Cited by volume:page number.]
- Dennett, Daniel (1993). "Review of Searle, *Rediscovery of the Mind*," *Journal of Philosophy* 60 (4): 193–205.
- Denis, Lara (ed.) (2010). *Kant's Metaphysics of Morals: A Critical Guide*. Cambridge: Cambridge University Press.
- Dietrichson, Paul (1962). "What Does Kant Mean by 'Acting from Duty'?" *Kant-Studien* 53 (1–4): 277–88.
- Dupré, Louis (1998). "Kant's Theory of History and Progress," *Review of Metaphysics* 52: 813–28.
- Dworkin, Ronald (2000). *Sovereign Virtue: The Theory and Practice of Equality*. Cambridge, MA: Harvard University Press.
- Formosa, Paul (2008). "All Politics Must Bend Its Knee Before Right': Kant on the Relation of Morals to Politics," *Social Theory and Practice* 34/2: 365–88.
- ____, Goldman, Avery, and Patrone, Tatiana (eds) (2014). *Politics and Technology in Kant*. Cardiff: University of Wales Press.

- Forster, Michael (2007). "Johann Gottfried Herder," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. URL: <http://plato.stanford.edu/entries/herder/> (accessed June 23, 2013).
- Galston, William (1975). *Kant and the Problem of History*. Chicago, IL: University of Chicago Press.
- Gazzaniga, Michael (2011). *Who's in Charge? Free Will and the Science of the Brain*. New York: Harper-Collins.
- Ginet, Carl (1980). "The Conditional Analysis of Freedom," in P. van Inwagen (ed.), *Time and Cause: Essays Presented to Richard Taylor*. Dordrecht: Kluwer.
- Goethe, Johann Wolfgang von (1989 [1832]). *Faust. Werke*, Hamburger Ausgabe, 14th edition, vol., 3, ed. Erich Trunz. Munich: C. H. Beck. [Cited by line number.]
- Gregor, Mary (1963). *Laws of Freedom*. New York: Barnes and Noble.
- Guyer, Paul (2000). *Kant on Freedom, Law and Happiness*. New York: Cambridge University Press.
- (2005). *Kant's System of Nature and Freedom: Selected Essays*. Oxford: Clarendon Press.
- (2006). *The Cambridge Companion to Kant and Modern Philosophy*. Cambridge: Cambridge University Press.
- Hahn, Songsuk, Susan (ed.) (2006). "Fichte's Intersubjective I," *Inquiry*, 49 (1). URL: <[http://www.tandfonline.com/\[ArticleDOI\]](http://www.tandfonline.com/[ArticleDOI])> (accessed June 21, 2013).
- Hausman, Daniel M. and Matt S. Waldren (2011). "Egalitarianism Reconsidered," *Journal of Moral Philosophy* 8: 567–86.
- Hayek, Frederick A. (1944). *The Road to Serfdom*. London: Routledge.
- (1976). *Law, Legislation and Liberty*. Chicago, IL: University of Chicago Press.
- Heidegger, Martin (1953 [1927]). *Sein und Zeit*. Tübingen: Mohr. Cf. (1962) *Being and Time*, tr. Edward Robinson and John Macquarrie, New York: Harper and Row.
- Herman, Barbara (1992). *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.
- Hill, Thomas, Jr (2002). *Human Welfare and Moral Worth*. Oxford: Oxford University Press.
- Honneth, Axel (1995). *The Struggle for Recognition: The Moral Grammar of Social Conflicts*, tr. Joel Anderson. Cambridge, MA: MIT Press.
- (2010). *The Pathologies of Individual Freedom: Hegel's Social Theory*, tr. Ladislaus Löb. Princeton: Princeton University Press.
- Hume, David (1967 [1739]). *A Treatise on Human Nature*, ed. Lewis Amherst Selby-Bigge. Oxford: Clarendon Press.
- Hursthouse, Rosalind (1999). *On Virtue Ethics*. Oxford: Oxford University Press.
- Husserl, Edmund (1960 [1929]). *Cartesian Meditations*, tr. Dorion Cairns. The Hague: Nijhoff.
- Hutcheson, Francis (1971 [1742]). *An Essay on the Nature and Conduct of the Passions and Affections. With Illustrations on the Moral Sense*, ed. Bernard Peach. Cambridge, MA: Harvard University Press.
- James, David (2011). *Fichte's Social and Political Philosophy: Property and Virtue*. Cambridge: Cambridge University Press.

- Joerden, Jan (1995). "From Anarchy to Republic: Kant's History of State Consistutions," in Hoke Robinson (ed.), *Proceedings of the Eighth International Kant Congress*, V 11. Milwaukee: Marquette University Press, pp. 139–56.
- Johnson, Robert (1996). "Kant's Conception of Merit," *Pacific Philosophical Quarterly* 77: 313–37.
- Jorgensen, J. and Newlands, S. (eds) (2009). *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*. Oxford: Oxford University Press.
- Kain, Patrick (2003). "Prudential Reasons in Kant's Anthropology," in Brian Jacobs and Patrick Kain (eds), *Essays on Kant's Anthropology*. Cambridge: Cambridge University Press, pp. 230–51.
- Kaufman, Alexander (1999). *Welfare in the Kantian State*. Oxford: Clarendon Press.
- Kaufmann, Walter (ed.) (1954). *The Portable Nietzsche*. New York: Viking.
- _____. (1956). *Existentialism from Dostoevsky to Sartre*. New York: Meridian.
- Kersting, Wolfgang (1992). "Politics, Freedom and Order," in P. Guyer (ed.), *The Cambridge Companion to Kant*. New York: Cambridge University Press, pp. 342–66.
- Kisilevsky, Sari and Stone, Martin (eds) (forthcoming). *Freedom and Force: Essays on Kant's Legal Philosophy*. Oxford: Hart Publishing.
- Kolodny, Niko (2005). "Why Be Rational?," *Mind* 114: 509–63.
- Korsgaard, Christine (2007). "The Normativity of Instrumental Reason," in Garrett Cullity and Beryls Gaut (eds), *Ethics and Practical Reason*. Oxford: Oxford University Press, pp. 215–54.
- _____. (2009). "The Activity of Reason," *Proceedings and Addresses of the American Philosophical Association* 83: 23–43.
- Ladd, John (ed.) (1973). *Ethical Relativism*. Belmont, CA: Wadsworth.
- Laitinen, Arto and Sandis, Konstantin (eds) (2010). *Hegel in Action*. Houndmills: Palgrave Macmillan.
- Locke, John (1823 [1689]). *Two Treatises of Government*, in *The Works of John Locke, A New Edition, Corrected*, Vol. V. London: Printed for Thomas Tegg; W. Sharpe and Son; G. Ofor; G. and J. Robinson; J. Evans and Co. Also R. Griffin and Co. Glasgow and J. Gumming, Dublin. [Cited as *Second Treatise*, by Chapter and Section (§) number.]
- _____. (1975 [1690]). *Essay Concerning Human Understanding*, ed. Peter Nidditch. Oxford: Oxford University Press.
- MacIntyre, Alison (2011). "Doctrine of Double Effect," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall edition). URL: <<http://plato.stanford.edu/archives/fall2011/entries/double-effect/>> (accessed June 23, 2013).
- Mack, Eric (1982). "In Defense of Blackmail," *Philosophical Studies* 41: 273–84.
- Mangan, Joseph (1949). "An Historical Analysis of the Principle of Double Effect," *Theological Studies* 10: 41–61.
- McGregor, Richard (2011). "Zhou's Cryptic Caution Lost in Translation," *Financial Times*, 10 June.
- Milgram, Stanley (1974). *Obedience to Authority: An Experimental View*. New York: Harper & Row.
- Mill, John Stuart (2001 [1862]). *Utilitarianism*, 2nd edition, ed. George Sher. Indianapolis: Hackett.
- Miller, David (1982). "Arguments for Equality," *Midwest Studies in Philosophy* 8: 73–83.

- _____. (1997). "Equality and Justice," *Ratio* 10: 222–37.
- Moyar, Dean (2011). "Kantian Right and the Realization of Free Agency," in Oliver Thorndike (ed.), *Rethinking Kant*, Vol. 3. Newcastle: Cambridge Scholars, pp. 124–57.
- Mueller-Vollmer, Kurt (1989). "Language Theory and the Art of Understanding," in Marshall Brown (ed.) *The Cambridge History of Literary Criticism*, Vol. 5. Cambridge: Cambridge University Press, pp. 162–84.
- Mulholland, Leslie (1990). *Kant's System of Rights*. New York: Columbia University Press.
- Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.
- _____. (1989). *The View from Nowhere*. New York: Oxford University Press.
- _____. (1993). "Moral Luck," in Daniel Statman (ed.), *Moral Luck*. Albany: State University of New York Press.
- Neuhouser, Frederick (1990). *Fichte's Theory of Subjectivity*. New York: Cambridge University Press.
- _____. (2008). *Rousseau's Theodicy of Self-Love*. Oxford: Oxford University Press.
- Nietzsche, Friedrich (1967). *The Will to Power*, tr. Walter Kaufmann and R. J. Hollingdale. New York: Random House.
- _____. (1974). *The Gay Science*, tr. Walter Kaufmann. New York: Random House.
- Nozick, Robert (1969). "Coercion," in Sidney Morgenbesser, Patrick Suppes, and Morton White (eds), *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*. New York: St. Martin's Press, pp. 440–72.
- _____. (1974). *Anarchy, State and Utopia*. New York: Basic Books.
- O'Neill, Martin (2008). "What Should Egalitarians Believe?," *Philosophy and Public Affairs* 36: 119–56.
- Parfit, Derek (2011). *On What Matters*. Oxford: Oxford University Press.
- Pettit, Philip (1997). *Republicanism: A Theory of Freedom and Government*. Oxford: Oxford University Press.
- Pippin, Robert (2006). "Mine and Thine: The Kantian State," in Guyer (ed.), *The Cambridge Companion to Kant and Modern Philosophy*. Cambridge: Cambridge University Press, pp. 416–46.
- Quante, Michael (2004). *Hegel's Concept of Action*, tr. Dean Moyar. Cambridge: Cambridge University Press.
- Ramachandran, V. S. (2004). *A Brief Tour of Human Consciousness*, New York: Pi Press.
- Rawls, John (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- _____. (2000). *Lectures on the History of Moral Philosophy*, ed. Barbara Herman. Cambridge, MA: Harvard University Press.
- Rendón, Carlos (ed.) (forthcoming). *Figure and Experience: Lessons about Hegel's Phenomenology of Mind*. Medellín.
- Ripstein, Arthur (2009). *Force and Freedom: Kant's Legal and Political Philosophy*. Cambridge, MA: Harvard University Press.
- Roemer, John (ed.) (1986). *Analytical Marxism*. Cambridge: Cambridge University Press.
- Rosen, Allen (1993). *Kant's Theory of Justice*. Ithaca, IL: Cornell University Press.
- Ross, W. D. (1930). *The Right and the Good*. Oxford: Clarendon Press.
- Rousseau, Jean-Jacques (1997 [1762]). *On the Social Contract and Other Later Political Writings*, ed. and tr. Victor Gourevitch. Cambridge: Cambridge University Press.

- Rush, Fred (ed.) (2013). *Yearbook for German Idealism*, vol. IX, Autumn. Berlin: Walter de Gruyter.
- Russell, Bertrand (1926). "Freedom in Society," *Harper's Magazine* (April).
- Sabini, John and Silver, Maury (2005). "Lack of Character? Situationism Critiqued," *Ethics* 115: 535–62.
- Sartre, Jean Paul (1956 [1943]). *Being and Nothingness*, tr. Hazel Barnes, New York: Philosophical Library.
- _____. (1964 [1938]). *Nausea*, tr. Lloyd Alexander. New York: New Directions.
- Satz, Debra (2010). *Why Some Things Should Not Be For Sale*. Oxford: Oxford University Press.
- Schmitt, Carl (1996 [1927]). *The Concept of the Political*, tr. George Schwab. Chicago, IL: University of Chicago Press.
- Scanlon, T. M. (1986). "Equality of Resources and Equality of Welfare: A Forced Marriage?," *Ethics* 97: 111–18.
- _____. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- _____. (2008). *Moral Dimensions: Permissibility, Meaning, Blame*. Cambridge, MA: Harvard University Press.
- _____. (forthcoming). "Reasons and Rationality," *Philosophy*.
- Schick, Theodore and Vaughn, Lewis (2002). *How to Think About Weird Things: Critical Thinking for a New Age*, 3rd edition. Boston, MA: McGraw-Hill.
- Schopenhauer, Arthur (1958). *The World as Will and Representation*, tr. E. J. F. Payne. New York: Dover Books. Cited by volume and section (§).
- Searle, John (1992). *Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- _____. (forthcoming), "Reasons and Rationality," *Philosophy*.
- Sen, Amartya (1992). *Inequality Reexamined*. Cambridge, MA: Harvard University Press.
- Setiya, Kieran (2007). *Reasons without Rationalism*. Princeton: Princeton University Press.
- Skinner, Quentin (1990). "The Republican Idea of Political Liberty," in Gisela Bock et al. (eds), *Machiavelli and Republicanism*. Cambridge: Cambridge University Press, pp. 293–309.
- Smith, Adam (2000 [1759]). *The Theory of Moral Sentiments*. Amherst, NY: Prometheus Books.
- Spinoza, Baruch (Benedict) (1982). *The Ethics and Selected Letters*, tr. Samuel Shirley. Indianapolis: Hackett. [Cited by Part, Proposition etc.]
- Stern, Paul (1986). "The Problem of History and the Temporality of Kant's Ethics," *Review of Metaphysics* 39: 505–45.
- Stern, Robert (2012). *Understanding Moral Obligation: Kant, Hegel, Kierkegaard*. Cambridge: Cambridge University Press.
- Stirner, Max (1995 [1844]). *The Ego and Its Own*, tr. David Leopold. Cambridge: Cambridge University Press.
- Temkin, Larry (1993). *Inequality*. Oxford: Oxford University Press.
- Thorndike, Oliver (2011). *Rethinking Kant*, Vol. 3. Newcastle: Cambridge Scholars.
- Timmons, Mark (ed.) (2002). *Kant's Metaphysics of Morals: Interpretive Essays*. Oxford: Oxford University Press.
- United Nations Universal Declaration of Human Rights (1950). URL: <<http://www.un.org/en/documents/udhr/index.shtml>> (accessed June 23, 2013).

- Weber, Max (1958 [1921]). "Politics as a Vocation," in H. H. Gerth and C. Wright Mills (eds), *From Max Weber: Essays in Sociology*. Oxford: Galaxy Books, pp. 77–128.
- Wertheimer, Alan (1999). *Exploitation*. Princeton: Princeton University Press.
- Wilcox, William H. (1987). "Egoists, Consequentialists, and Their Friends," *Philosophy and Public Affairs* 16: 73–84.
- Williams, Bernard (1981). *Moral Luck*. Cambridge: Cambridge University Press.
- Wolff, Jonathan (2002). *Why Read Marx Today?* Oxford: Oxford University Press.
- Wood, Allen (1990). *Hegel's Ethical Thought*. New York: Cambridge University Press.
- ____ (1995). "Exploitation," *Social Philosophy and Policy* 12 (2): 136–58.
- ____ (1997). "Exploitation," in Kai Nielsen and Robert Ware (eds.), *Exploitation*. New York: Humanities Press, 1997, pp. 1–22.
- ____ (1999). *Kant's Ethical Thought*. New York: Cambridge University Press.
- ____ (2002). *Unsettling Obligations: Essays on Reason, Reality and the Ethics of Belief*. Stanford: CSLI Publications.
- ____ (2004). *Karl Marx*, 2nd edition. London: Taylor and Francis.
- ____ (2008). *Kantian Ethics*. New York: Cambridge University Press.
- Zwolinski, Matt (2008). "The Ethics of Price Gouging," *Business Ethics Quarterly* 18 (3): 347–78,

Index

- A Few Good Men*, 115–116
A priori, 44, 48, 53, 57, 100, 104, 106, 132, 195, 215–217
 Abortion, 277, 307–308
 Achelis, H. N., 164
 Adams, R. M., 121, 139–143
 Adler, M., 84
 Advertising, 291–294
 Alexander the Great, 251
 Allison, H., 27–30
 Alternatives, acceptable and unacceptable, 278–282
 America, 90n, 95n, 96n, 103n, 108n, 116n, 141–142, 175n, 252, 269, 270n, 272, 272n, 295, 301, 303–305, 307–314
 Anarchism, 12
 Anscombe, G. E. M., 239
 Apartheid, fall of, 272n
 Aquinas, Thomas, 243–245
 Arendt, H., 136n
 Aristotle, 4, 64–65, 129, 148–151, 153–154, 158, 287–288
 Arneson, R., 253, 270n
Ataraxia, 315–317
 Autonomy, 11, 68–69, 72, 79, 166, 173, 209, 277, 293; Formula of, 79; See also Self-constraint

 Barbarism, 87, 98, 124, 306
 Baron, M., 16, 18, 146, 283, 286–292
 Bauer, B., 260–261
 Beauvoir, S., 316
 Bakunin, M., 12
 Beccaria, C., 271–272
 Beck, L. W., 120
 Behaviorism, 198
 Beiser, F., 120, 131, 214–215
 Ben Ali, Z., 106
 Benedict, R., 128
 Bentham, J., 129, 145, 157–159, 256
 Berlin, I., 131, 132
 Bernstein, A., 98
 Blackmail, 297
 Blanc, L., 259–260, 271
 Bolshevism, 314–315
 Bonaparte, Napoleon, 251
 Bondage, bonded labor, 96, 227, 284–286
 Boyd, R., 151, 159
 Broome, J., 41, 50
 Bush, G. W., 96n, 103n, 108n, 111n, 116n
 Buss, S., 288
 Byrd, B. S., 84

 Caesar, Julius, 250
 Capitalism, 4, 7, 11, 39, 84, 88–89, 149, 154, 227, 252–273, 297–302, 311–314
 Cartesian, see Descartes, R.
 Cather, W., 271
 Character, see Virtue; weaknesses of, 287, 289–290
 Charity, faith based, 247–248
 Choice, see Coercion, Freedom, Will; tragic, 116, 282
 Christianity, 14, 122, 244, 259, 261
 Churchland, P., 211
 Citizen, passive, 88–89
 Civil disobedience, 97–99
 Classes, 253–267
 Coercion, 8–12, 38, 64, 70–88, 94–99, 109, 115, 134, 177, 207, 209, 260–261, 269, 274–286, 294–303; “worse off relative to a baseline,” 281
 Cognizance (*Kenntnis*), 240–241
 Cohen, G. A., 253, 257, 258, 270n
 Cohen, H., 84
 Collins, R., 7
 Communication, freedom of, 6, 66
 Community, 5–7, 10–12, 67–69, 220, 222
 Comprehension (*Begreifen*), 182–185
 Conditional analysis of “could have done otherwise,” 171–172
 Confucian ethics, virtues for rulers, 303–304
 Connell, F. J., 243–246
 Conscience, 14, 33, 96, 98, 178–179, 234, 251, 280, 308
 Consequentialism, 4, 144–163
 Constant, B., 113
 Constraint, 278; see also Self-constraint, Coercion
 Conviction, 182–185
 Communism, see Socialism
 Cultural determinism, 128; see also Relativism

 Damasio, A., 195
 Darwall, S., 42
 Darwin, C., 130
 Death penalty, 271–272
 Deception, 286, 289
 Declaration (*Aussage, declaratio*), see Lying, Kant on
 Dennett, D., 196

- Deontology, 144–163
 Descartes, R., 6, 168, 186, 192, 194–199, 210, 215, 247
 Desire, 4, 27, 35, 40, 43–45, 50, 51–56, 60–62, 76–77, 124, 133, 150–156, 171–172, 204, 209–210, 218–225, 235–238, 275, 280, 294; see also Inclination
 Despotism, 9, 90, 95–98, 109, 125, 272, 306
 Diderot, D., 44, 196
 Dietrichson, P., 18
 Dignity (*Würde*), 32–33, 67–69, 87, 109, 134, 136n, 158, 175, 269, 298, 301
 Dogmatism, 173–176, 180–181
Dolus directus, *dolus indirectus*, 236
 Domination, freedom from, 224–227, 294–296, 302; see also Right
 Dominion of thoughts (*Gedankenherrschaft*) (Stirner, Marx), 265–266
 Dostoevsky, F., 109
 Double effect, 243–246
 Drugs, 308–309
 Dupré, L. 136n
 Duty (*Pflicht*), 3–4, 10, 13–38, 40–42, 46, 51, 58–65, 70–72, 77–85, 98, 107, 112–113, 150n, 164–165, 177, 232–233, 243–244, 248–250, 280, 282, 284, 308–309; see also Self-constraint; Virtue, duties of
 Dworkin, R., 253
 Ebbinghaus, J., 84
 Education, 6, 135, 208–211, 222, 225, 253, 257, 260, 263, 290, 293–294, 296, 313
 Egoism, 264–266; theoretical, 200
 Egypt, Ancient, 124, 127
 Emotion, see Feeling
 End (*Zweck*), 4, 72–78, 80–81, 145–152, 206, 220
 Ends, realm of (*Reich der Zwecke*), 67, 79, 154, 222; Formula of, 79
 Engels, F., 106, 253–255, 260, 314
 Enlightenment (*Aufklärung*), 44–45, 66, 98, 119–121, 125–143, 222, 272
 Epictetus, 225
 Epicurus, 192
 Equality, 84–85, 165, 252–273; procedural, 255
 Equity, right of, 97
 Esteem (*Hochschätzung*), 20
 Ethical Life (*Sittlichkeit*) (Hegel), 4–5, 229–230, 235, 250
 Ethics, 279–80; history of, 14; see also, Morality, Right
 Ethics of disposition (*Gesinnungsethik*) (Weber), 93–95
 Ethics of responsibility (*Verantwortungsethik*) (Weber), 93–95, 110, 115–116
 Eudaimonism, 124; see also Happiness
 Evidentialism, 140–141, 308
 Exploitation, 275–277, 296–303
 Externality, 293
 Facticity, 170
 Faith, in freedom, 173–176, 190–191; historical, 137–143
 Feeling, 14, 36, 42, 45, 55, 64, 120, 124–126, 161, 170–172, 182, 186, 194, 198, 200, 242, 248, 263, 268, 279–280, 287
 Fichte, J. G., 1, 2–3, 4–12, 42, 67, 89, 119, 146–152, 154, 164–193, 195, 199–228, 229, 247, 255–256, 269, 294–295, 306–307, 315–316
 Firearms, our human right to freedom from privately owned, 310
 First person standpoint, 189–190, 211–212
 Flikschuh, K., 71
 Føllesdal, D., 200n
 Formosa, P., 117
 Formulas of the Moral Law, see Moral Law, Formulas of
 Forster, M., 120, 132
 Forsyte' change, 158
 Foucault, M., 5
 Freedom, 1–12, 72–73, 274–314, absolute, 3, 164–193; external, see Right; formal and material, 172–173; in self-regarding matters, 307–309; of the will, 164–193, 230–231, 249; of political expression, 310–311; subjective, 229–231, 239–240, 246–249, 294; transcendental, 165–167; free development of each, 259–260, 266, 314, 316; see also Will, Coercion, Manipulation, Right
 French Revolution, see Revolution, French
 Freud, S., 5, 168
 Friedrich Wilhelm II, King of Prussia, 117
 Frost, D., 96n
 Galston, W., 136n
 Galsworthy, J., 158
 Garve, C., 107
 Gazzaniga, M., 189
 Geertz, C., 128
 General will, 88, 99–108
 Ginét, C., 172
 Goethe, J. W., 315–316
 Good will, 15, 22, 31–37, 175, 233–234
 Good, highest (*summum bonum*), 61, 154–155
 Good, practical, 61–63
 Grand Inquisitor, 109
 Greed, 287
 Greece, Ancient, 124–125
 Gregor, M., 84
 Guyer, P., 71, 77n, 84
 Hamann, J. G., 120
 Happiness, 14, 21, 32–36, 46–47, 61–63, 65–69, 74–81, 84–86, 108–109, 124–129, 133–136,

- 151–159; 230, 240, 302; see also Reason, prudential
- Hausman, D., 253
- Hayek, F. A., 84, 87n
- Health care, 38, 86, 117
- Hedonism, 157–158; see also Pleasure
- Hegel, G. W. F., 1–2, 4–5, 8–10, 199, 214–251, 262, 269, 294, 305, 316
- Heidegger, M., 150, 170n
- Herder, J. G., 1, 7, 119–137, 305–306, 315–316
- Herman, B., 145, 156
- Herskovits, M., 130
- Hierarchy (Stirner), 265
- Hill, T., 48, 57
- History, see Faith, historical; stages of, 124–127, see also *Humanität* (Herder); “verdict of,” 111
- Hobbes, T., 44, 63, 196
- Holiday, B., 289–290
- Honneth, A., 214
- Honor, 37
- Hruschka, J., 84
- Humanität* (Herder), 131–132, 138
- Humanity (*Menschheit*), 2, 17, 19, 59, 68–69, 72–74, 75, 79, 80–81, 87, 101, 116, 124–126, 131, 269, 298–299; friend of, 24, 35–36; formula of, 74, 79, 101; right of, 113–115, 284
- Humboldt, W., 121, 210
- Hume, D., 35, 48, 63, 122, 123, 210
- Hurka, T., 103
- Hursthouse, R., 116
- Husserl, E., 199–201
- Hutcheson, F., 14, 35, 41, 120
- I, the (*das Ich*), 166–171, 203–228; essence of, 167–170; individuality of, 203–205; as will, 170–173
- Idealism, 173–176, 180–181
- Ideology, 84–85, 91, 227, 265–266, 272, 314
- Iglesias, D., 116n
- Ilting, K. H., 241
- Imagination, 53, 58, 59, 141, 171, 173, 182–188
- Impeachment, 99
- Imperatives, 45–63
- Imputability (*Zurechnung*), 4–5, 17, 229–251; see also Responsibility
- Inclination (*Neigung*), 4, 11, 16–28, 34–37, 40, 43–45, 49, 51–62, 65, 74, 77, 124, 133, 171, 280; see also Desire
- Incorporation thesis, 27–30
- Individualism, 5–7, 69, 229, 266
- Inequality, 84–89, 252–253
- Injustice, general, 85–89, 284
- Innocence, 37–39
- Intention (*Absicht*), 235–239
- Intersubjectivity, 5–8, 42, 65–69, 194–213, 220–228
- Intuitions, moral, 144–145
- Iraq, U.S. invasion of, 108n, 111n, 141, 272
- James, D., 89
- Joerden, J., 98
- Johnson, R., 17–18
- Judaism, 260–261
- Justice, distributive, 255–259
- Justification, 41–43, 67–69; externalist, 188–189
- Kain, P., 54n
- Kant, I., 1–4, 8–12, 13–118, 123, 132–143, 231–235, 149–150, 154, 156, 165–167, 173, 178, 186, 192–193, 200–201, 214, 229, 238, 248, 255–256, 272n, 276, 280, 284, 294–295, 298, 306, 308, 315–316
- Kaufman, A., 84
- Kennickell, A., 252n
- Kersting, W., 84
- Kolodny, N., 41, 50
- Korsgaard, C., 41, 60
- Labor, see Bondage, Capitalism, Wage Labor, Labor Unions, Slavery
- Labor Unions, 298, 300–301
- Lametttrie, J. O., 196
- Lassalle, F., 300
- Law, universal, 16, 34–37, 60–62, 70–71, 78–80; see also Universal Law, Formula of
- Leibniz, G. W., 167, 210
- Levinas, E., 199
- Libertarianism, 83–89
- Locke, J., 55–56, 195, 210, 294–295
- Louis XIV, King of France, 127
- Love, 289; of humanity, 19; see also, Feeling, Sympathy, Humanity, friend of
- Luck egalitarianism, 270n
- Ludwig, B., 71
- Luke, St., 259
- Lying, see Deception; Kant on, 111–114, 232, 276, 284
- MacIntyre, A., 243, 245
- Mack, E., 297
- MacPherson, J., 122
- Madoff, B., 161
- Maimon, S., 184
- Mandeville, B., 44
- Mangan, J., 245
- Manipulation, 286–294
- Marcus Aurelius, 225
- Market, “free,” 227, 252–253, 292–294, 311–314; see also Capitalism
- Marx, K., 1–2, 5, 7, 8–10, 106, 154, 214, 226, 252–273, 298–302, 304, 306, 314–316
- Marxism, 142; see also Marx, K., Stalinism, Bolshevism

- Master-servant (*Herr-Knecht*), 8, 224–225
Mendelssohn, M., 123
Merit (*Verdienst*), 15–20, 231–234, 250; see also, Imputability
Metaethics, 120, 172, 268, 277
Mill, J. S., 146–148, 152–155, 159–161, 315
Miller, D., 253
Moral law, Formulas of, 16, 34–37, 74, 79, 101
Moral luck, 243–249
Moral worth, 13–20, 31–33
Moralist, despotic, 95; political, 92; valet, 251
Morality, 3–5, 177–181; *Moral, Sitten* (Kant), 70–73, 91–95; *Moralität* (Hegel), 229–251; as class ideology (Marx), 262–266; see also Moral Law, Formulas of
Moralized concepts, 275–277, 288–289
Moses, 69
Motive, 21–22, 27–34, 43, 80, 238–239
Moyar, D., 73
Mubarak, H., 106
Mueller-Vollmer, K., 210
Mulholland, L., 84
Murder, as *homicidium dolosum*, 275–276
- Nagel, T., 2, 246
Natorp, P., 84
Nazism, 98, 109, 121–122, 272
Necessity, right of, 97
Negligence, 239–243
Neuhouser, F., 182, 226–227
Nietzsche, F., 5, 158, 186–187, 227
Nixon, R., 96n
Novalis (Friedrich Hardenberg), 210
Nozick, R., 84, 281, 311, 312n
- Obama, B., 272n
Oedipus, 235, 240
Omnipotence, paradox of, 285
O'Neill, M., 253
Orient, ancient, 124–125
Ossian, 122
Owen, R., 300
Owl of Minerva, 262
- Parfit, D., 45, 156, 190
Paul, St., 244
Pettit, P., 146, 160, 294
Pippin, R., 73
Phoenicians, ancient, 124
Pleasure, 14, 36, 38, 55, 87, 126, 135, 141, 154–158; see also Hedonism
Politician, moral, 92
Politics, 90–118, 253–260
Post-modernism, 121
Praise, 20; see also Esteem
Pressuring, 281, 286–287, 289
Price-gouging, 297
- Progress, belief in, see Faith, historical
Proudhon, P., 300
Providence, divine, 131–136; see also Faith, historical
Property, 11, 85, 88, 96, 100, 108, 124, 222, 230, 255–256, 261, 263, 265, 276, 297, 300, 310
Publicity, principles of, 9, 99–108
Purpose (*Vorsatz*), 235–239
Pyrrho, 120
- Quine, W. V. O., 190
- Racism, 91, 116n, 303–304
Ramachandran, V.S., 186–187
Rationalism, 63–65, 122–123, 173, 186, 195, 289–290
Rationality, requirements of, 40–42; see also Reason, Reasons
Rawls, J., 14, 62, 255, 258
Reagan, R., 90n, 252
Realpolitik, 91–93, 108–118
Reason, 40–69; instrumental, 47–52; moral, 60–62; prudential, 52–60, 63–67; theoretical and practical, 48–51, 63, 77n, 137–140, 143, 173, 181–182
Reasons, 3–7, 18, 27, 130, 140–141, 144–150, 155–158, 159n, 162–163, 174, 176, 185–187, 189–191, 201, 208–209, 213, 221–222, 238–239, 255, 276, 279–282, 290, 304; *pro tanto*, 40–45, 50–52, 58, 61–62
Rebellion, 97–99, 105–107
Recall by election, 99
Recognition (*Anerkennung*), 7–10, 65–69, 214–228
Regret, agent, 249
Reinhold, K. L., 165, 216
Relativism, 128–132
Religion, 117, 121, 122, 127, 134, 192, 228, 260–261, 263, 265, 299, 306–309, 315, 316
Respect (*Achtung*), 175, 196, 222, 226, 280, 291; for the moral law, 22–23, 34, 117, 280; for humanity in persons, 12, 32–33, 64, 67–68, 75, 79, 158–159, 299, 309; self-respect, 253; see also Recognition, Dignity
Responsibility, moral, 177–178, 277–278, 285–286; *Schuld sein an*, 235–236; see also Imputability
Revolution, French, 98, 222, 253–254, 272; right of, see Rebellion
Rice, C., 96n
Right (*Recht*), 7–12, 70–90, 93–118, 253–260, 279–80, 284; law of, 70–71, 82; principle of, 70–73; of intention, 237; of knowledge, 235; “the right thing to do,” 280
Ripstein, A., 84, 98
Rome, ancient, 124
Roosevelt, F., 109–110n

- Rosen, A., 84
Rousseau, J.-J., 44, 85, 226–227, 255–256, 282–283, 294–295
Rozeboom, G., 57
Russell, B., 298
- Salads*, 175n
Sartre, J. P., 150, 167, 175n, 184, 191, 214
Satz, D., 87n, 284, 293
Scanlon, T. M., 50–51, 157–158, 238–239, 253
Scheuermann, W., 109–110n
Schick, T., 211
Schiller, F., 4, 211
Schlegel, A.W., 210
Schleiermacher, F. D., 121
Schmitt, C., 92, 94
Schopenhauer, A., 166
Searle, J., 196
Self-certainty (*Selbstgewissheit*), 223–225
Self-constraint, 12, 21–40, 40–47, 149, 280; see also Duty
Self-sufficiency (*Selbständigkeit*), 167–170
Sen, A., 25 September 11, 2001, 141–142
Serfdom, road to, 87n
Setiya, K., 63
Sex, 95, 307–309, 311
Shaftesbury, Anthony Ashley Cooper, Third Earl of, 14
Shakespeare, W., 122
Shame, 303–305
Shapshay, S., 110n
Shopping, 293–294
Skilling, J., 161
Skinner, Q., 294
Slavery, 96, 126, 134, 225, 255, 257–258, 265, 269, 270n, 271–272, 298, 303–304, 308n; spiritual, 91, 179–181, 316; reason as slave of the passions, 48–49, 152–153
Smith, A., 42, 44, 293–294
Socialism, 83–89, 121, 254, 258–259, 262, 268, 271, 273, 300, 313; see also Marxism, Soviet Union
Solidity (*Gediegenheit*), 235
Solipsism, 199–203
Soviet Union, invasions of Hungary and Czechoslovakia, 272n; fall of, 272n
Species, human, 3, 110, 124, 131, 133–139, 154, 209, 218–220, 261, 266, 270–273, 299, 301
Spinoza, B., 123, 164, 166–168, 175n, 178, 180–183, 186–188
Stalinism, 314
Stein, E., 199
Stephani, H., 165
Stern, P., 136n
Stern, R., 4
Stirner, M., 265–267
- Striving (*Streben*), 25–26, 85, 121, 126, 131, 137–138, 154, 167–168, 204, 212, 219–220, 250, 315
Sturm und Drang, 120
Subjectivity, moral (Hegel), see Freedom, subjective
Summons (*Aufforderung*), 6–7, 178, 205–210, 178, 221–223
Sumner, W. G., 129–130
Sympathy, 14, 16, 19–20, 24, 33, 35–38, 197, 287; see also Feeling
- Tarski, A., 172
Temkin, L., 253
Tillich, P., 139
Time, 171, 182–185; see also Imagination
Transcendental philosophy, 5–6, 102, 168–170, 180–181, 187, 192–193, 198–213, 214–226
- Understanding (*Verstand*), 182–187; maxims of, 136–137
Unique (*Einzig*), 265
United Nations General Declaration of Human Rights, 86–89
Universal law, Formula of, 34–37, 79, 116; see also Law, universal
Unsociable sociability (*ungesellige Geselligkeit*), 38, 267
Utilitarianism, 66, 68, 84, 86, 109, 124, 129, 146, 152
- Vaughn, L., 211
Virtue (*Tugend*), 4, 11, 13–14, 23–26, 29–38, 45, 47, 72, 87, 94, 122–129, 140–141, 146, 159, 164–165, 174, 177, 280, 303–304; duties of, 24, 29–30, 32, 47, 77
Virtue theories of practical reason, 63–65
Voltaire, 84, 123
Vorländer, K., 84
Voter suppression, 116n
- Wage labor, 88–89; see also Capitalism
Waldren, M., 253
War, preventive, 107–108; right of, 241
Wavering (*Schweben*), 184, 187–188; see also Imagination
Way of thinking (*Denkungsart*) (Herder), 121–132
Weber, M., 92–95, 110, 115–116
Weinrib, J., 95
Weinrib, E., 106
Weisshuhn, F. A., 164
Wertheimer, A., 275
Wilcox, W., 147
Will, 170–171, 204; holy, 34–35; see also Reason, Freedom of the will; I, the; General Will; Good will

- Willaschek, M., 71
- Williams, B., 147, 249
- Williams, H., 103
- Wittgenstein, L., 190, 199
- Wolff, J., 260
- Wöllner, J. C., 117
- Wood, A., 44, 60, 84, 89, 111, 141, 192, 232, 242,
256, 260, 284, 296
- World historical individuals, 250
- Worldhood, 170
- Xenophobia, 303
- “You can’t handle the truth,”
115–116
- Zhou enlai, 272
- Zwolinski, M., 297

